



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

ODSTRANĚNÍ ZNÁMÉHO SIGNÁLU Z NAHRÁVKY

REMOVAL OF A KNOWN SIGNAL FROM A RECORDING

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. RADOMÍR URBÁNEK

VEDOUCÍ PRÁCE

SUPERVISOR

Doc. Dr. Ing. ČERNOCKÝ JAN,

BRNO 2018

Abstrakt

Cílem této diplomové práce je návrh a implementace metody, která ze zvukové nahrávky dokáže odstranit známý signál. V úvodní části práce jsou popsány vlastnosti zvuku a jeho šíření prostorem, zpracování signálu v počítači a hledání impulsní odezvy systému. Dále je popsán postupný vývoj metody vedoucí k odstranění známého signálu z nahrávky. Následuje popis a implementace této metody. Dále je popsáno, jakým způsobem a na jakých datech bude metoda testována. Na závěr je testování vyhodnoceno, navrženo zlepšení a další možné pokračování práce.

Abstract

The goal of this diploma thesis is to design and implement method for removing known signal from recorded sound. In the introductory part of the thesis are described the properties of sound and its propagation through the space, digital signal processing on the computer and the computing of the system impulse response. The sequential development of the methods leading to the removal of the known signal from the recording is described further. The following part contains a description and implementation of this method. It also describes how and at what data the method will be tested. Finally, testing is evaluated, improvements and further possible work is proposed.

Klíčová slova

Odstranění šumu, zpracování signálu, systém, impulsní odezva, analýza řeči, frekvenční doména

Keywords

Disturbance removal, signal processing, system, impulse response, speech analysis, frequency domain

Citace

URBÁNEK, Radomír. *Odstranění známého signálu z nahrávky*. Brno, 2018. Diplomová práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Doc. Dr. Ing. Černocký Jan,

Odstranění známého signálu z nahrávky

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana docenta Jana Černockého. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

.....
Radomír Urbánek
23. května 2018

Poděkování

Děkuji vedoucímu mé práce, panu docentu Janu Černockému za vedené mé práce, připomínky a odborné konzultace, které mi pomohly k pochopení problematiky týkající se tohoto tématu. Dále děkuji panu doktoru Szőkemu za pomoc při sběru dat pro testování a vyhodnocení.

Obsah

| | | |
|----------|---|-----------|
| 1 | Úvod | 3 |
| 2 | Odstranění známého zvuku z nahrávky | 5 |
| 2.1 | Zvuk a jeho šíření prostorem | 5 |
| 2.2 | Zpracování signálu | 6 |
| 2.3 | Konvoluce | 9 |
| 2.4 | Korelace | 11 |
| 2.5 | Identifikace systému | 11 |
| 2.6 | Spektrální analýza | 15 |
| 2.7 | Způsoby vyhodnocení kvality zvukových signálů | 15 |
| 3 | Data | 17 |
| 3.1 | Data určená pro testování | 17 |
| 3.2 | Simulovaná data | 18 |
| 3.3 | Reálná data | 22 |
| 4 | Vývoj metody | 24 |
| 4.1 | Útlum zvuku v prostředí | 24 |
| 4.2 | Synchronizace signálů | 25 |
| 4.3 | Výběr částí signálu pro výpočet impulsní odezvy | 26 |
| 4.4 | Výpočet impulsní odezvy | 27 |
| 5 | Návrh metody | 30 |
| 5.1 | Synchronizace | 30 |
| 5.2 | Identifikace systému | 30 |
| 5.3 | Aplikace na ruchový signál | 32 |
| 5.4 | Odečtení od vstupního signálu | 32 |
| 6 | Implementace | 34 |
| 6.1 | Odstranění známého signálu z nahrávky | 34 |
| 6.2 | Lokace signálu | 36 |
| 6.3 | Výpočet ruchového signálu | 36 |
| 6.4 | Odečítání ve spektru | 39 |
| 6.5 | Spuštění programu | 39 |
| 7 | Vyhodnocení a testování | 40 |
| 7.1 | Vyhodnocení simulovaných dat | 41 |
| 7.2 | Vyhodnocení reálných dat | 45 |

| | |
|-------------------|-----------|
| 8 Závěr | 47 |
| Literatura | 49 |

Kapitola 1

Úvod

Spousta zvukových dat obsahuje rušení, které dělá užitečný zvukový obsah méně srozumitelným, nebo zcela nesrozumitelným. Zjistit informace z takových dat je pak velmi obtížné jak pro lidského posluchače, tak pro případné automatické zpracování. Takové rušení je však možné odstranit. Existuje několik technik, které to mohou zvládnout s uspokojivým výsledkem. Tato problematika však není stále dobře vyřešená a neexistuje zcela univerzální postup a je třeba vybrat metodu s ohledem na vstupní data.

Tato práce se zabývá takovými zvukovými nahrávkami, které obsahují mluvené slovo rušené různými zdroji, například hrajícím televizorem nebo rádiem. Toto rušení je problémem především při v automatickém dolování informací z řeči. Rušení může být jak zcela náhodný signál, tak zvuk, jehož obsah známe nebo můžeme získat. Tato práce se zabývá odstraněním právě takového rušení. To ovšem situaci příliš nezlehčuje. Rušivý signál totiž nemusí být zcela stejný, jako je zaznamenaný v rušeném signálu. Zvláště v případě zvukové nahrávky je rušivý zvuk ovlivňován různými faktory. Při průchodu zvukových vln v prostředí dochází k zeslabení, v případě místnosti odrazu zvuku od stěn. Zvukové vlny se pak skládají, vznikají různé interference a ozvěny. Zkreslovat nahraný zvuk ale nemusí jen prostředí. Každé elektronické zařízení jako mikrofony, reproduktory či zesilovače zvuk různými způsoby zkresluje a není tak možné od nahraného zvuku rušivý obsah odečíst.

Ukázkovým příkladem mohou být odposlouchávací zařízení, která používá například policie pro odposlechy. Tato zařízení typicky umístěná v místnosti při nahrávání požadované řeči zachycují i zvuky, které jsou rušivé a mohou způsobit nesrozumitelnost zachycené konverzace. Přitom oním rušivým elementem nemusí být pouze náhodný zvuk z okolí, ale i hrající televizor či rádio. Případně záměrně umístěná rušička. V takovém případě je možné rušící zvuk dohledat v archivech, případně simulovat stejným přístrojem.

Dalším příkladem může být problém při mistrovství světa ve fotbale v roce 2010, které se pořádalo v Jihoafrické republice. Afričtí fanoušci při zápasech hráli na svůj domorodý nástroj zvaný vuvuzela. Jeho specifický zvuk se však příliš nelíbil televizním stanicím, které se ho snažili ve svých vysíláních utlumit. Obdobný problém nastává u hlasem ovládaných zařízení, která přitom sama zvuk vydávají, jako jsou například televize či rádio. Zvuky, které tato zařízení vydávají, se nahrávají na jejich mikrofon sloužící pro příjem hlasových povelů od uživatelů. Jelikož je mikrofon u takového zařízení velice blízko reproduktorům, analýza takových povelů rušených vlastním zvukem může být bez odstranění rušení jen těžko řešitelným problémem.

Mojí osobní motivací pro tuto práci bylo studium a získání praktických zkušeností z problematiky zpracování reálných signálů. Tato oblast se stává s pokročilejší automatizací v průmyslu i běžném životě stále významnější. Přitom její řešení nejsou jednoduchá a

nástroje jako jsou neuronové sítě a umělá inteligence obecně nemusí být vždy úspěšné či v praxi použitelné. Algoritmy a postupy nezvládají pracovat obecně a musí být adaptovány na konkrétní problematiku.

V kapitole 2 je popsán teoretický úvod do problematiky, fyzikální podstata zvuku a jeho šíření a faktory, které zvukové signály ovlivňují a bude je třeba brát v úvahu v dalších částech práce. Dále je zde uvedeno základní zpracování zvukového signálu v počítači. Také uvádí přehled vybraných technik, které se používají při získávání informací z především jednorozměrných signálů. Kapitola 3 pak popisuje data, která byla vytvořena pro implementaci, testování a vyhodnocení navržené metody. V části 4 je uveden komplexní rozbor problematiky a možných řešení. Další kapitola 5 prezentuje navrženou metodu pro odstranění známého signálu z nahrávky. Následující kapitola 6 popisuje implementaci navržené metody v prostředí Matlab. V kapitola 7 je implementovaná metoda otestována a hodnotí její úspěšnost. Poslední kapitola 8 obsahuje vyhodnocení metody, srovnání s jinými možnostmi řešení a navrhuje další možná rozšíření práce.

Kapitola 2

Odstranění známého zvuku z nahrávky

Představme si následující situaci. Člověk, mluvčí, hovoří před mikrofonom, který jeho promluvu nahrává a vytváří tak zvukový signál. V pozadí v blízkosti mluvčího přehrává zařízení jiný, zvukový signál. Tento signál ve formě zvukových vln prostupuje prostředím a je rovněž zaznamenán mikrofonom. Možné vlivy prostředí jsou popsány v kapitole 2.1. Signály se v mikrofonu smíchají a na výstupu je jeden signál, který obsahuje promluvu člověka rušenou zvukem ze zařízení. Tato situace je ukázána na blokovém schématu 2.1.

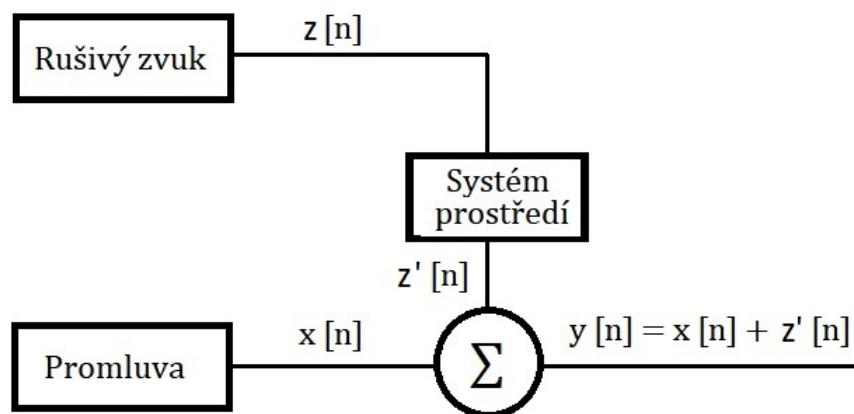
Úkolem je získat pouze signál obsahující promluvu člověka při znalosti signálu, kterým byla promluva rušena. Řešení ovšem musí být takové, aby nevyžadovalo další zásahy člověka a probíhalo v krátkém čase, nejlépe v reálném čase tak, aby bylo možné rušení odstraňovat přímo v průběhu nahrávání. Na výstupu zařízení je tak signál, který se zpracovává v počítači, případně jiném elektronickém zařízení. Techniky, které pro zpracování signálu a jeho analýzu slouží, jsou teoreticky popsány v sekci 2.2.

V této kapitole jsou následně popsány další techniky, které slouží pro analýzu signálu či jsou nápomocné pro jejich další zpracovávání či dolování informací. Mohou být specifické pro danou úlohu, nebo jen její část.

2.1 Zvuk a jeho šíření prostorem

Z fyzikálního pohledu je zvuk vlnění hmotných částic šířících se v prostředí. Rychlost šíření zvuku je konstantní a závislá na struktuře nosiče. Rychlost zvuku ve vzduchu je 345 metrů za sekundu v místnosti o pokojové teplotě [16]. Pokud zvukové vlny narazí na pevnou překážku, část energie zvuku je pohlcena, část projde skrz a část se odrazí. Odražené vlny mohou interferovat s původními vlnami ze zdroje a mohou je tak zesílit, nebo naopak zeslabit, případně vyrušit úplně. Tomuto jevu se říká reverberace. Různá prostředí produkují různé formy reverberace v závislosti na pozici a vzdálenosti překážek.

V této práci se budeme nadále věnovat šíření zvuku o frekvenci slyšitelné člověkem, tedy v rozmezí přibližně 20 Hz až 20 kHz, šířící se v uzavřené místnosti. V takovém případě se zvuk šířený od zdroje odrazí od kolmých stěn místnosti (naznačeno na obrázku 2.2). Detektor zvuku, kterým může být člověk, případně mikrofón zaznamenávající zvuk, tak detekuje jak zvukové vlny přímo od zdroje, tak také zvukové vlny odražené od stěn místnosti a ovlivněné útlumem a vzájemnými interferencemi. Každá místnost tak má svou



Obrázek 2.1: Blokové schéma signálu zachyceného mikrofonom. $x[n]$ signál promluvy, $z[n]$ signál rušení, $z'[n]$ signál rušení po průchodu prostředím, $y[n]$ výsledný signál

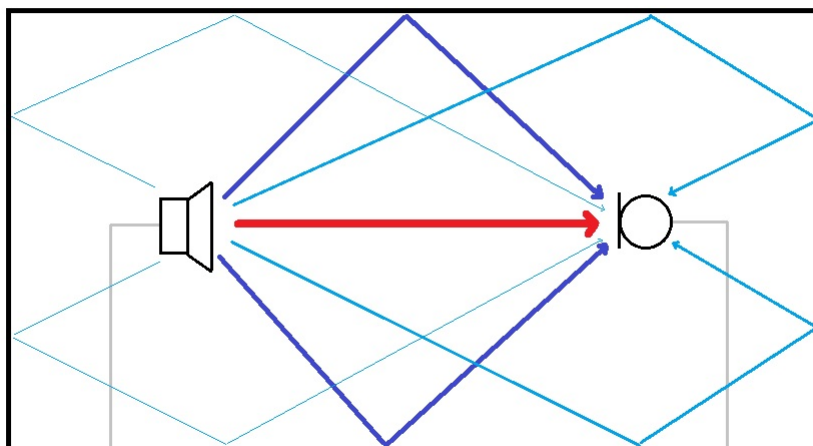
charakteristiku, která lze popsat impulsní odezvou. Typická impulsní odezva místnosti je zobrazena na obrázku 2.3.

Při průchodu zvuku prostředím je třeba brát v úvahu následující faktory [3]:

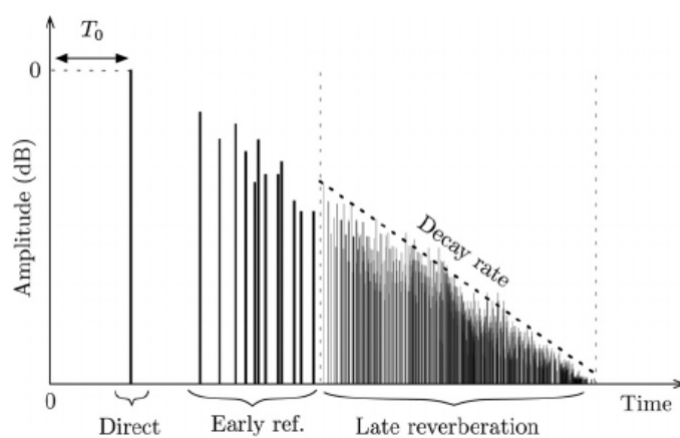
- Útlum zvuku - se vzdáleností od zdroje se snižuje intenzita zvuku. Tento útlum se skládá z více vlivů. Nejvýznamnějším je rozdělování energie na vzrůstající plochu. Útlum způsobuje také absorpce vzduchu nebo vliv teploty. Jako útlum je možné označit také absorpci překážek v místnosti. Útlum je přitom rozdílný pro jiné frekvence zvuku.
- Odraz zvuku - k odrazu zvuku dochází při dopadu zvukových vln na pevnou překážku. Přitom je část energie vlny překážkou pohlcena. Amplituda odražené vlny je menší a její fáze posunuta. Úhel dopadu vlny se rovná úhlu odrazu.
- Difuzní odraz - je způsoben drsností povrchu a nerovností povrchu
- Difrakce - difrakce nastává, když zvuková vlna dopadne na okraj překážky. Díky difrakci se zvukové vlny mohou dostat i do míst geometrického stínu [3]. Difrakční jevy lze rozdělit do tří skupin, difrakce na šterbině, hraně a překážce [2] (zobrazeno na obrázku 2.4).

2.2 Zpracování signálu

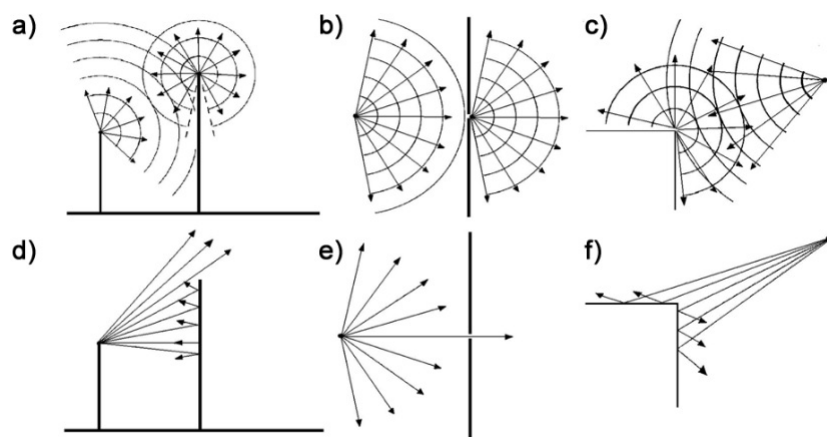
Zpracování zvuku není příliš vhodné provádět na celém vstupním signálu. To platí zvláště v případě signálů obsahujících zvukovou stopu. To už z důvodu, že může být zpracováván průběžně. Proto se typicky dělí na úseky, kterým říkáme rámce. Délka rámců by měla být zvolená tak, aby dostatečně malá, aby mohl být signál na daném úseku pokládán za stacionární, ale přitom dostatečně velká, aby bylo možné dostatečně přesně odhadnout požadované parametry [4]. Rámce se přitom mohou i překrývat, takže stejná část signálu



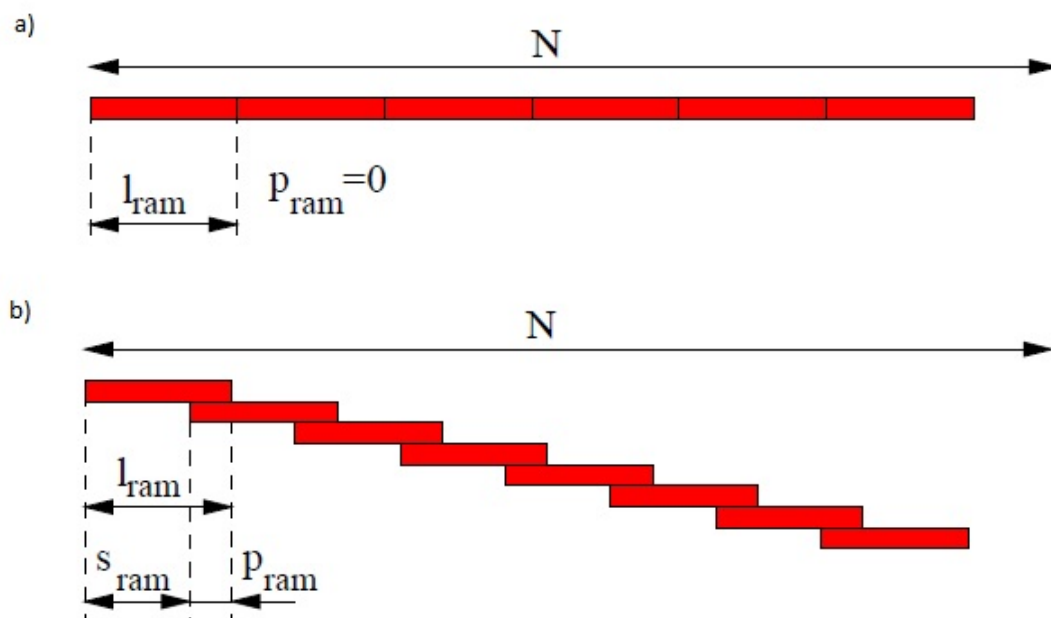
Obrázek 2.2: Šíření zvuku od zdroje k cíli buď přímo (červená) nebo odrazy od stěn místnosti (modrá).



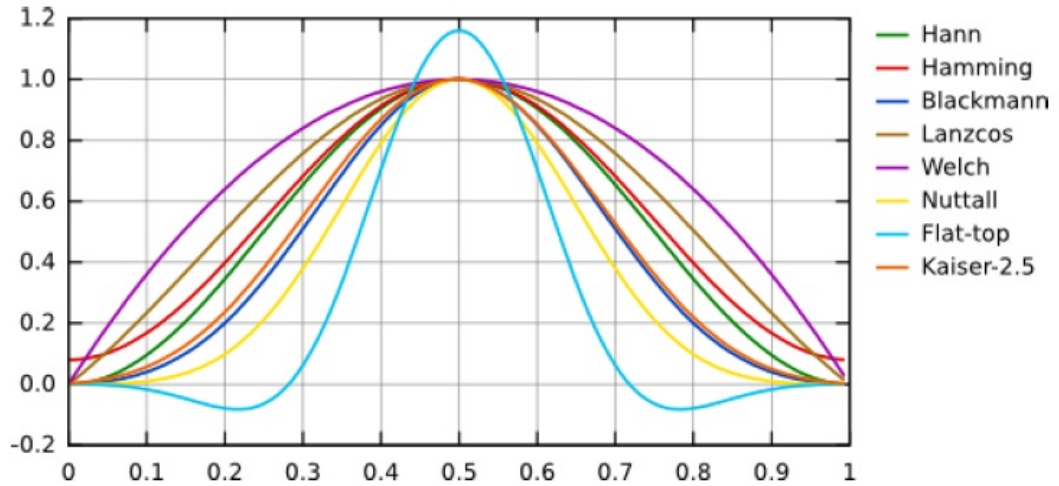
Obrázek 2.3: Příklad impulsní odezvy uzavřené místnosti (převzato z [16]).



Obrázek 2.4: Difrakce zvukových vln na šterbině, hraně a překážce při nižších frekvencích (obrázky a, b c) a při vyšších frekvencích (d, e, f) (převzato z [2]).



Obrázek 2.5: Rámce za a) s překrytím, za b) bez překrytí (Převzato z [4])



Obrázek 2.6: Vybrané průběhy okenních funkcí (Převzato z [8])

může být obsažena ve dvou rámcích. Překrytí se provádí pro plynulejší přechod parametrů, ovšem zvyšuje výpočetní nároky [4]. Překrývání rámců je zobrazeno na obrázku 2.5

Rámec se ze signálu získá operací zvanou windowing. Tato operace násobením předem definovaného signálu zvaného okno vybere pouze tu část, kterou požadujeme [14]. Nejjednodušší okenní funkce je pravoúhlé, či obdélníkové okno. To je definováno následovně:

$$w[n] = \begin{cases} 1 & \text{pro } 0 \leq n \leq l-1 \\ 0 & \text{jinde} \end{cases} \quad (2.1)$$

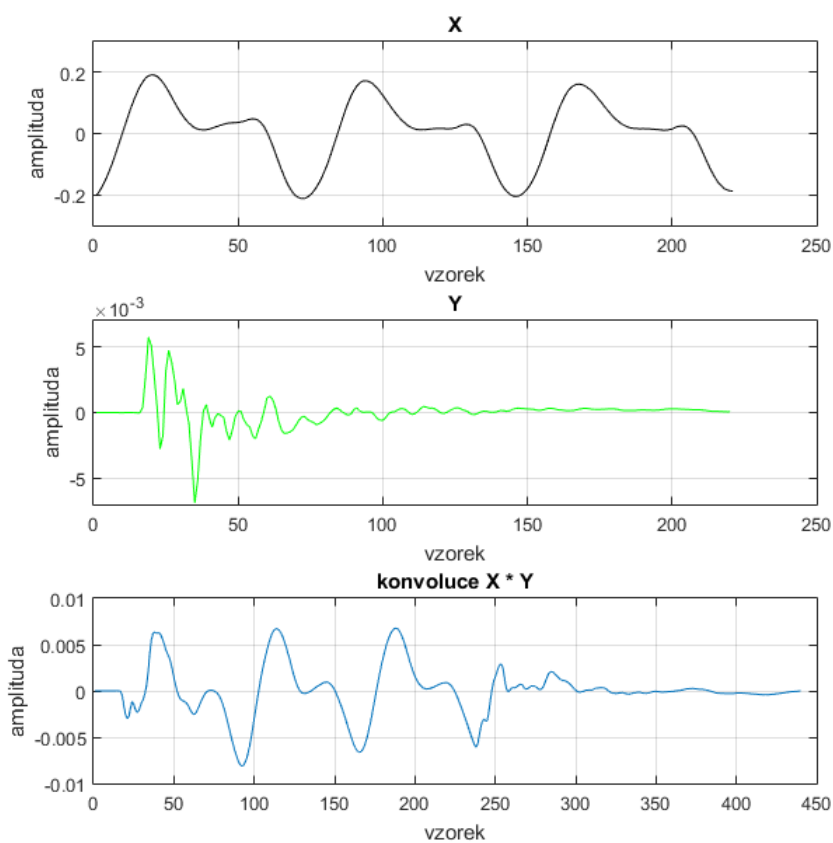
kde l je délka rámce. Často je užitečné na okrajích vybrané části signálu hodnoty utlumit. Používá se celá řada okenních funkcí. Vybrané z nich jsou ukázány na obrázku 2.6.

2.3 Konvoluce

Konvoluce je matematická operace dnes známá především pro její využití v počítačové grafice při zpracování dvojrozměrných signálů, tedy obrazu pomocí konvolučního jádra. U jednorozměrných signálů se používá například při zjištění impulsní charakteristiky. Konvoluce dvou diskretních signálů má vzorec podobu následující:

$$(f * g)([n]) = \sum_{m=-\infty}^{\infty} f[m]g[n-m], \quad (2.2)$$

kde f a g jsou konvoluované signály, a n aktuálně zpracovávaný vzorek. Příklad konvoluce dvou signálů je zobrazen na obrázku 2.7. Na prvním grafu je periodický signál x vyňatý ze zvukového záznamu. Na druhém je impulsní odezva y reálné místnosti. Na posledním je jejich konvoluce.



Obrázek 2.7: Konvoluce dvou signálů. Signál x (černě) je krátký úsek zvukového signálu, y (zeleně) je impulsní odezva místnosti. Modře je vyznačena jejich konvoluce.

2.4 Korelace

Velmi podobnou operací k operaci konvoluce je korelace. Ve zpracování signálu je korelace operátor, který určuje vzájemnou podobnost tvaru dvou signálů. Někdy se také nazývá Cross-Correlation Function (CFF). Běžně se tento operátor používá pro nalezení periodicity signálu nebo zpoždění dvou stejných nebo či podobných signálů. V této práci využijeme především její schopností nalézt v neznámém signálu známý vzor. Výsledkem korelace jsou korelační koeficienty, které udávají podobnost vstupních signálů. Korelace dvou diskretních signálů f a g je dána vztahem:

$$R[k] = \sum_{n=-\infty}^{\infty} f[n]g[k+n] \quad (2.3)$$

Jak je patrné, vzorec je velmi podobný se vzorcem diskretní konvoluce 2.2. Rozdíl je ve směru, v jakém jsou vzorky signálu g zpracovávány. Pokud je signál korelován se sebou samým, mluví se o autokorelaci. Tato metoda je ve zpracování signálu vhodná pro určení posunutí, neboli lagu. Samotná CFF ovšem selhává v případě, kdy je velký rozdíl mezi energiemi porovnávaných signálů [5]. Pro takové případy je vhodné použít Normalizovanou Cross-Correlation Function (NCCF). Ta je dána vztahem

$$NCCF(m) = \frac{\sum_{n=mr}^{mr+N-1} s(n)s(n-m)}{\sqrt{E_f E_g}} \quad (2.4)$$

Jde tedy CFF, kde E_f a E_g jsou energie obou funkcí. Energie E funkce f se v diskretním prostoru vypočítá podle vzorce:

$$E = \sum_{n=-\infty}^{\infty} f^2(n), \quad (2.5)$$

kde n je zpracovávaný vzorek. Na obrázku 2.8 je znázorněn příklad autokorelace, kdy je periodický signál X krátkého zvukového záznamu korelován se sebou samým, však posunutým o několik vzorků Y . Výsledkem jsou autokorelační koeficienty, ze kterých je možné vypočítat zpoždění.

2.5 Identifikace systému

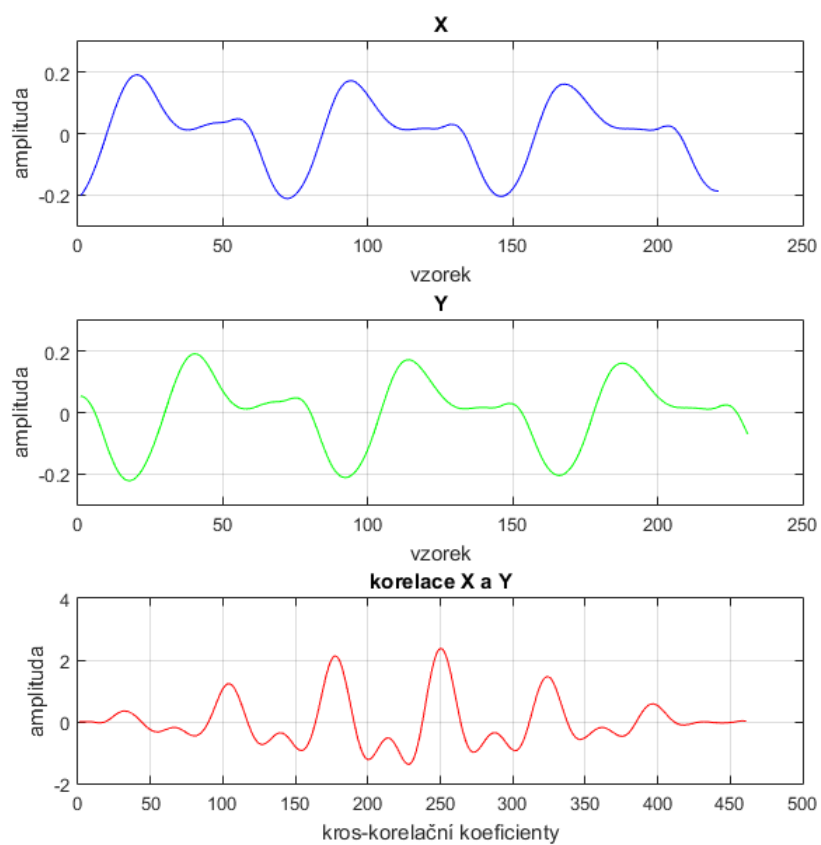
Termínem identifikace systému budeme v této práci označovat proces, při kterém je na základě dat ze vstupu systému a jeho výstupu určeny parametry odpovídající impulsní odezvě $h[n]$. Přitom budeme brát v úvahu pouze lineární dynamické systémy. Problém identifikace je najít operátor G a případně i spektrum přdaného rušení v [10].

Systém může být identifikován ve dvou doménách, časové a frekvenční. Doménu určuje typ dat, na jakých je systém identifikován. Pro časovou oblast jsou potřeba data v známé formě, kdy každý vzorek odpovídá danému času určeného časovou periodou, tedy převrácené frekvenci.

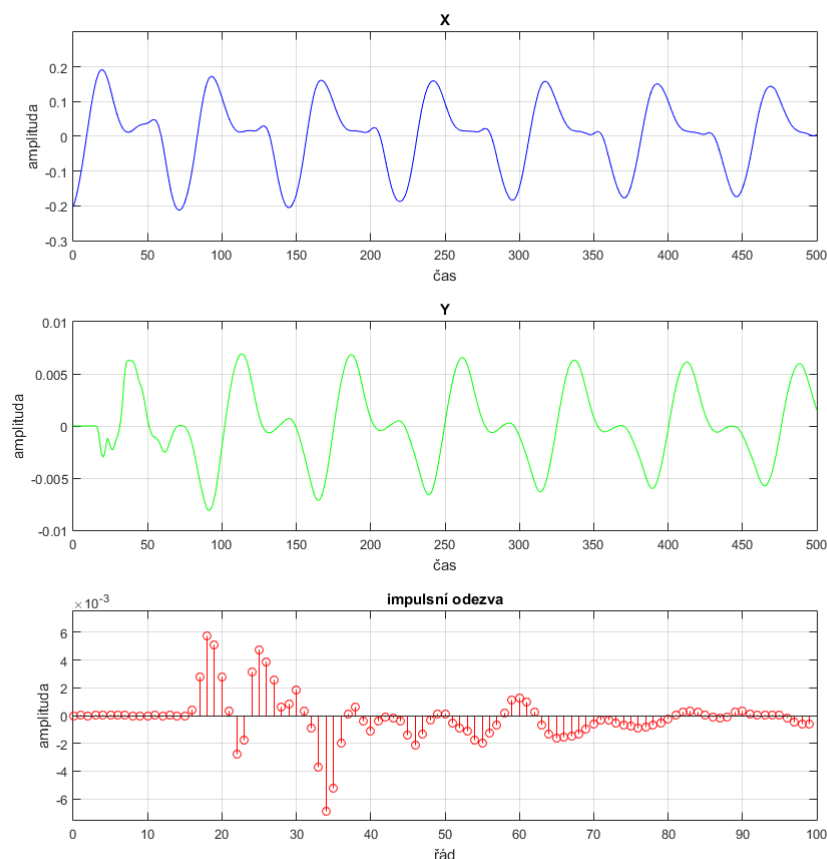
Při práci v diskretním čase v časové oblasti h je pak možné zjistit ze vztahu

$$y[k] = h[l] * x[k], \quad (2.6)$$

kde x a y jsou diskretní signály a symbol $*$ značí operátor konvoluce popsané v sekci 2.3. Délka impulsní odezvy může být jiná než vstupního signálu. Ve frekvenční oblasti



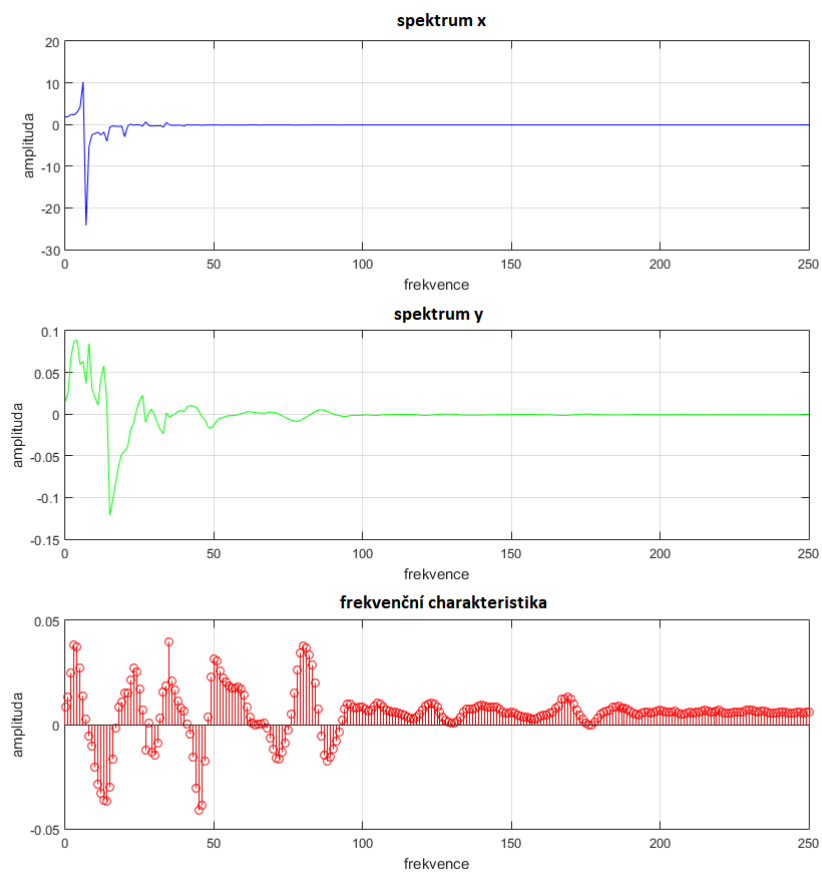
Obrázek 2.8: Korelace dvou signálů, kde signál y (červeně) je zpožděný o 20 vzorků proti signálu x (modře). Červeně je vyznačeny jejich kros-korelační koeficienty.



Obrázek 2.9: Odhad systému mezi dvěma signály v časové oblasti. Modře je zobrazen původní signál, zeleně signál po průchodu místností. Červeně je zobrazena vypočtená impulsní odezva místnosti.

$$y[k] = h[k]x[k], \quad (2.7)$$

kde x a y jsou diskrétní signály. Délka frekvenční charakteristiky h musí být stejná, jako je vstupní signál. Výsledný signál je tak dán pouze vynásobením vzorků vstupního a výstupního signálu. Složitost násobení dvou vektorů je výrazně nižší, než operátor konvoluce. Proto je i výpočet impulsní odezvy systému ve frekvenční oblasti výrazně rychlejší, než v časové [9]. Grafy na obrázku 2.9 zobrazují vypočtenou impulsní odezvu mezi dvěma signály v časové oblasti. Na prvním grafu je krátký úsek zvukového signálu, na druhém stejný úsek signálu po průchodu místností. Na třetím grafu je pak na základě těchto signálů vypočtená impulsní odezva místnosti. Stejný případ pak zobrazují grafy na obrázku 2.10, ovšem ve frekvenční oblasti. Signály z obrázku 2.9 jsou převedeny pomocí Fourierovy transformace popsané v sekci 2.6 do frekvenční domény a až poté je z nich vypočtena impulsní odezva podle vzorce 2.7.



Obrázek 2.10: Odhad systému mezi dvěma signály ve frekvenční oblasti. Vstupní signály odpovídají spektřům signálů na obrázku 2.9 a shodný je i systém, na kterém je odhad prováděn. Zobrazena je první polovina spektra.

2.6 Spektrální analýza

Při získávání informací ze signálu, který je složen z více harmonických složek, může být vhodné provést spektrální analýzu. Takovou charakteristiku mají téměř všechny reálné signály. Podstatou spektrální analýzy je rozložit signál na jednotlivé harmonické složky a určit jejich zastoupení v daném signálu. Pro zápis harmonických složek se využívá komplexních čísel, aby byla obsažena i informace o fázi. O signálu převedeného na jeho spektrální podobu říkáme, že je ve frekvenční oblasti. Pro reprezentaci signálu ve frekvenční oblasti použijeme Fourierovu transformaci [4]. Fourierova transformace je proces, při kterém je zadaný signál aproximován harmonickými signály. Výsledkem jsou komplexní koeficienty Fourierovy řady. Pro práci s digitálními signály je určena Diskrétní Fourierova Transformace DFT. Ta je dána vztahem:

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi}{N}kn}, \quad (2.8)$$

kde $k = [0, N - 1]$. Výsledkem je diskrétní frekvenční spektrum, kde hodnota prvku DFT $X[k]$ odpovídá zastoupení frekvence [5]:

$$\frac{k}{N}F_s, \quad (2.9)$$

kde F_s vřezkovácí frekvence a N je celkový počet vzorků [13]. Jeden vzorek N tak odpovídá jedné určité frekvenci. Fourierova transformace je však výpočetně náročná operace. Pro praktické výpočty se tak používá efektivnější metoda výpočtu DFT zvaná FFT (z anglického Fast Fourier Transform). Jde o velmi efektivní metodu, kdy je používáno dílčích výsledků místo jednotlivého výpočtu pro každou hodnotu v signálu. Metoda je velmi efektivní [15],[1]. Složitost DFT je $N\log_2 N$, ale FFT pouze $N\log_2 N$. FFT pracuje nejrychleji se signály o délce 2^N vzorků.

Fourierova transformace je reverzibilní operace. Z koeficientů Fourierovy transformace je tedy možné zpětně dopočítat hodnoty signálu v čase [6]. Této operaci se říká inverzní Fourierova transformace. Opět má svoji efektivnější variantu zvanou IFFT (Inverse Fast Fourier Transform).

2.7 Způsoby vyhodnocení kvality zvukových signálů

Vyhodnocení kvality zvukových dat je problematická záležitost a neexistuje jedna technika, která by zaručovala úspěšné srovnání. Metody dělíme na objektivní, kde kvalitu dokážeme vyhodnotit algoritmem, a subjektivní, kde je kvalita hodnocena lidskými posluchači [4]. Pokud jsou data určena pro posluchače, nikoliv automatické zpracování, je složité najít hodnotící kritérium, které by lidského posluchače reprezentovalo. Vhodným řešením je tedy do hodnocení zapojit budoucí uživatele. Často používanou technikou zejména v síťových aplikacích nebo testování nových kodeků je získání parametru MOS. Tento parametr lze získat subjektivní i objektivními metodami. Obvykle však kvalitu hodnotí skupina posluchačů a výsledky řadí do pětibodové stupnice:

- 1 - unacceptable (nepřijatelné)
- 2 - poor (slabé)

- 3 - fair (slušné)
- 4 - good (dobré)
- 5 - excellent (výborné)

Jedná se ale o časově a finančně náročnou metodu [17]. Jako další subjektivní metody si uvedeme například DRT (Diagnostic Rhytme Test) a DAM (Diagnostic Acceptability Measure). Ovšem i tyto metody mohou být pro dané použití příliš náročné. Z tohoto důvodu se přistupuje k objektivním metodám, jako je například SNR (Signal to Noise Ratio), tedy odstup signálu od šumu:

$$SNR = 10 \log_{10} \left\{ \frac{\sum_{n=0}^{N-1} s^2(n)}{\sum_{n=0}^{N-1} [s(n) - \hat{s}(n)]^2} \right\}, \quad (2.10)$$

kde s je signál a \hat{s} šum. Výsledek je v decibelech. Jako další metodu si uvedeme keprstrální vzdálenost (CD z Cepstral Distance), která hodnotí signály ze znalostí referenčního signálu a je tak vhodná pro simulovaná data. Počítá se z keprstrálních koeficientů a vyhodnocuje Euklidovou vzdáleností. Mezi další objektivní techniky patří SRMR (Speech-to-reverberation modulation energy ratio) vzdálenost nebo LLR (Log Likelihood Ratio).

Kapitola 3

Data

V rámci vývoje implementace metody byla vytvořena data, která se využívala na testování programu či jeho jednotlivých součástí. K vývoji byla využívána data simulovaná v počítači, kde byl záznam zvuku obsahující mluvené slovo ovlivněn tak, aby připomínal zvuk po průchodu místností rušený rádiem. Pro tento účel se využívaly dva soubory:

- prednaska.wav - záznam přednášky obsahující mluvené slovo (délka 100 sekund) použitý jako užitečný signál
- ruseni.wav - záznam nahraný z rádia, obsahující nejdříve hudbu, poté mluvené slovo (délka 50 sekund) použitý jako rušení

Soubor s užitečným signálem byl rušen různým způsobem tak, aby postupně simuloval všechny vlivy popsané v kapitole 2 s různým stupněm intenzity.

3.1 Data určená pro testování

Pro testování metody byly vytvořeny dva soubory dat. Prvním byla data simulovaná v počítači, kde byl záznam signálu obsahující mluvené slovo ovlivněn tak, aby připomínal nahraný zvuk po průchodem místností zašumněný rádiem. Druhým byla data nahraná v reálné místnosti. Užitečný zvuk byl pouštěn z reproduktoru a nahráván mikrofonom. Z jiného reproduktoru znělo rušení. Pro oba testovací soubory dat byl použit stejný užitečný signál, na kterém promlouvalo několik mluvčích v nahrávací místnosti tak, aby na něm nebyl patrný žádný šum. U všech tří souborů dat bylo shodné i známé rušení. To bylo rozděleno do čtyř kategorií podle předpokládané obtížnosti:

- klasická hudba - bez hlasu (soubor 01_allegro.wav), délka 294 sekund.
- rocková hudba - melodie a zpěv (soubor 02_rock.wav), délka 225 sekund.
- rap - více hlasu (soubor 03_rap.wav), délka 297 sekund.
- mluvené slovo - bez melodie, obsahuje promluvu člověka podobně jako užitečný signál (soubor 04_zpravy.wav), délka 307 sekund.

Je očekávané, že rušení obsahující hlas lidí půjde odstranit obtížněji, než rušení obsahující pouze melodii. Nejobtížnější na odstranění by mělo být mluvené slovo, kde bude nutné odlišení mezi dvěma řečníky.

Soubor s užitečným signálem (*studenti.wav*) má délku 332 sekund.



Obrázek 3.1: Místnost auly, ze které byla získána impulsní odezva pro vytvoření simulovaných dat (získáno z [7]).

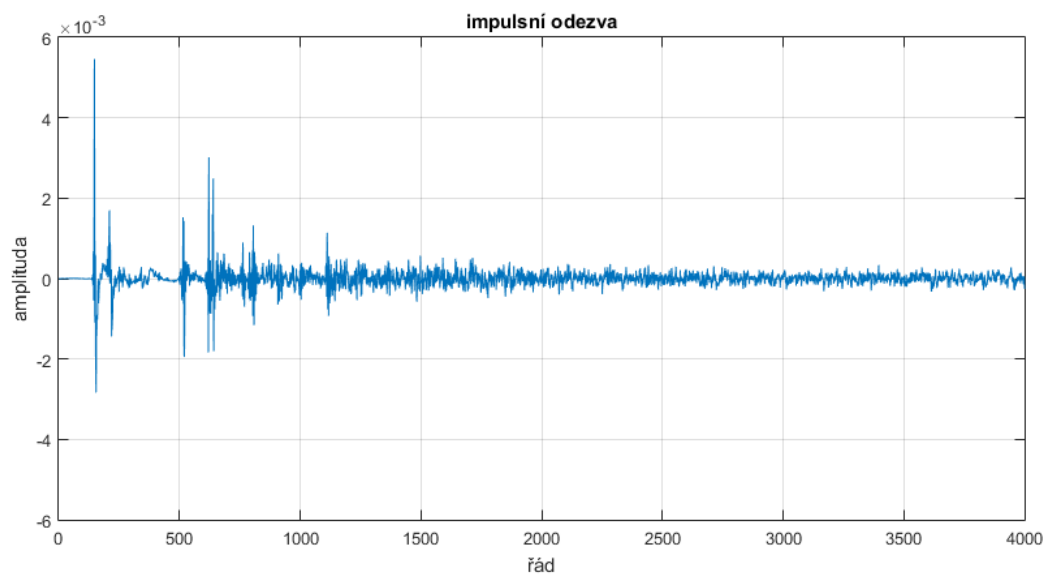
3.2 Simulovaná data

Jelikož nebylo možné provést rozsáhlé vyhodnocení ve velkém počtu místností a zvukových záznamů, byla vytvořena simulovaná data, která nahrazovala rozsáhlejší testování. Další přínos je, že mohla být vyhodnocena metodou SNR 2.10. Nahrávky obsahující užitečný signál byly stejně jako nahrávky obsahující rušení navzorkovány na vzorkovací frekvenci 16 kHz . Nižší frekvence byla vybrána pro rychlejší vyhodnocení. Každá z nahrávek pak byla vynásobena koeficientem, který označíme za faktor zesílení. Tyto koeficienty byly čtyři (0.2, 0.5, 1 a 2) a měly simulovat rozdíl vzdálenosti a hlasitosti zdroje rušícího zvuku od mikrofону. Vzniklo tak 16 různých nahrávek z rušením, které byly přimíchány k užitečnému signálu.

Data byla vytvořena pomocí knihovny impulsních odezev, která byla získána z Aachen University [7]. Z této knihovny byly vybrány impulsní odezvy čtyřech rozdílných místností:

- aula - velká místnost bez překážek (místnost zobrazena na obrázku 3.1 a její impulsní odezva na obrázku 3.2)
- chodba - dlouhá úzká místnost (místnost zobrazena na obr. 3.3 a její impulsní odezva na obrázku 3.4)
- přednášková místnost - střední místnost s překážkami (místnost zobrazena na obr. 3.5 a její impulsní odezva na obrázku 3.6)
- zasedací místnost - malá místnost (místnost zobrazena na obr. 3.7 a její impulsní odezva na obrázku 3.8)

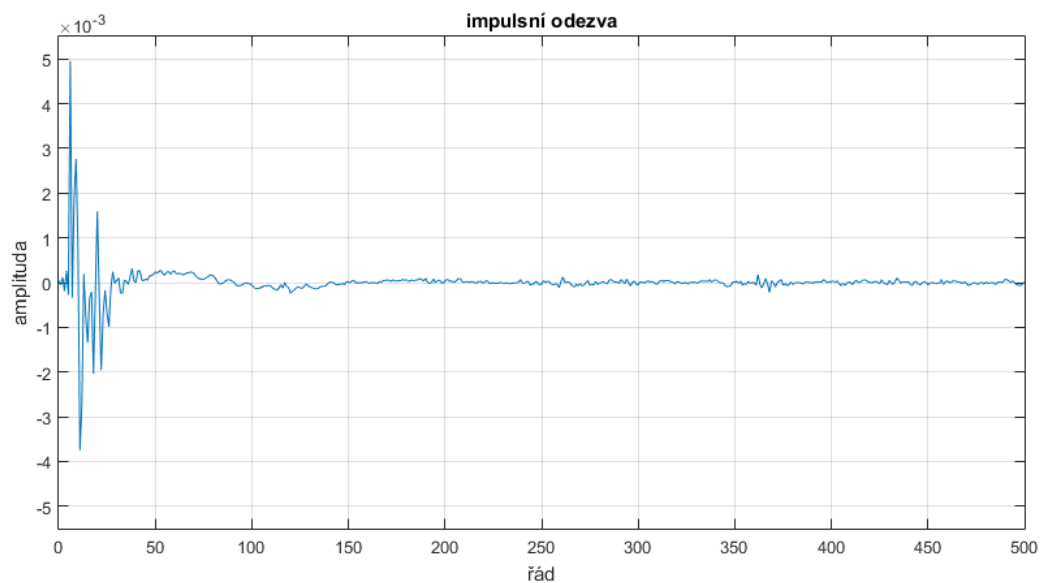
Každá z impulsních odezev pak byla aplikována na všechny vytvořené nahrávky pro simulaci průchodu místností. Dohromady tak bylo vytvořeno 64 simulovaných nahrávek k otestování. Všechny tyto nahrávky měly délku a vzorkovací frekvenci shodnou s původní, užitečnou nahrávkou.



Obrázek 3.2: Impulsní odezva auly.



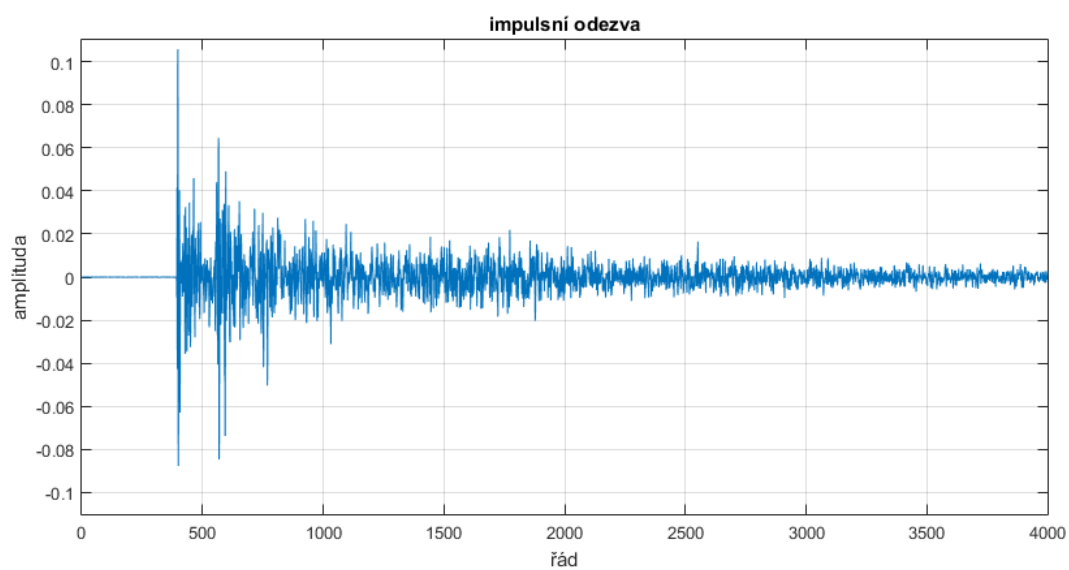
Obrázek 3.3: Místnost chodby, ze které byla získána impulsní odezva pro vytvoření simulovaných dat (získáno z [7]).



Obrázek 3.4: Místnost chodby, ze které byla získána impulsní odezva pro vytvoření simulovaných dat (získáno z [7]).



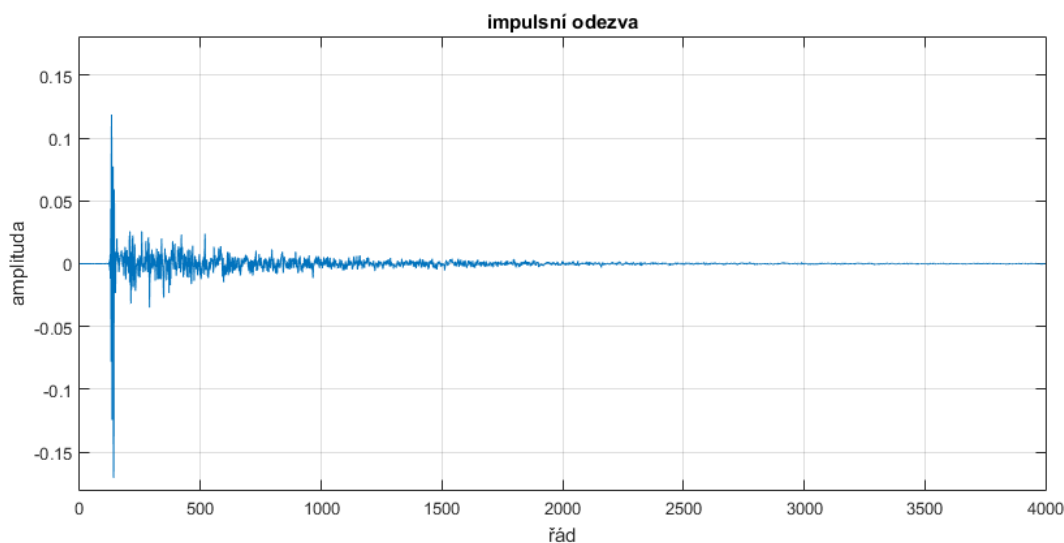
Obrázek 3.5: Přednášková místnost, ze které byla získána impulsní odezva pro vytvoření simulovaných dat (získáno z [7]).



Obrázek 3.6: Impulsní odezva přednáškové místnosti.



Obrázek 3.7: Zasedací místnost, ze které byla získána impulsní odezva pro vytvoření simulovaných dat (získáno z [7]).

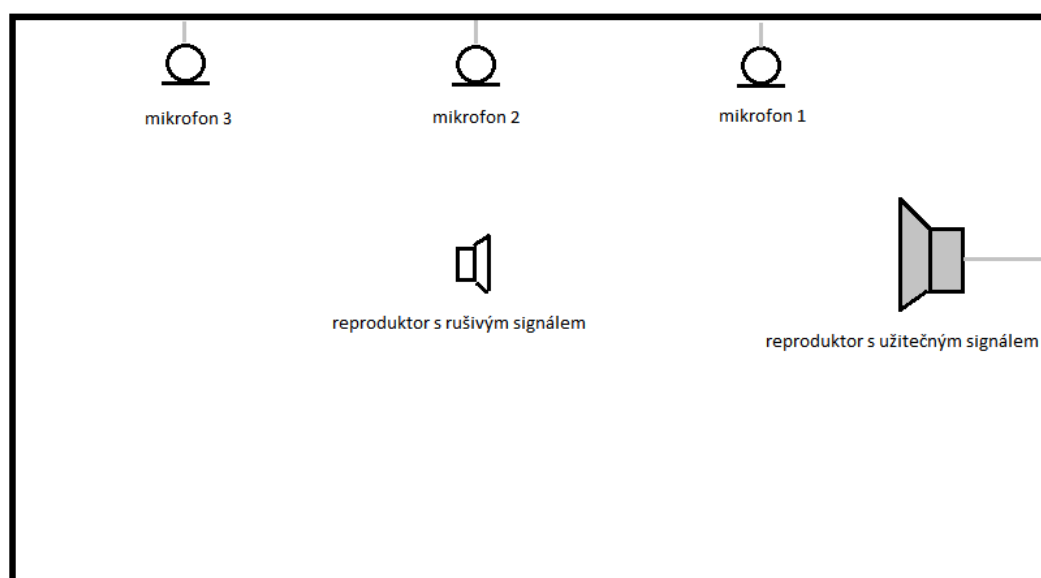


Obrázek 3.8: Impulsní odezva zasedací místnosti.

3.3 Reálná data

Druhým souborem dat byly nahrávky vytvořené v reálném prostředí. Zdrojem užitečného signálu byl kvalitní reproduktor, zdrojem rušení reproduktor notebooku. Zdroj zvuku byl na jedné straně místnosti a zdroj rušení uprostřed. Zvuk byl zaznamenáván na tři mikrofony v rozdílných vzdálenostech od zdroje postavené tak, aby první mikrofon byl nejblíže užitečnému signálu, druhý mikrofon uprostřed místnosti blízko zdroje rušení, a třetí na druhé straně místnosti. Situace je naznačena na schématu 3.9. Místnost měla velikost podobnou jako zasedací místnost zobrazená na obrázku 3.7.

Opět byly použité stejné čtyři rušivé záznamy popsané v úvodu. Také vzorkovací frekvence byla ponechána na 16 kHz pro rychlejší vyhodnocení. Nahrávání probíhalo způsobem, kdy se nejdříve spustí nahrávka s užitečným signálem, a několik vteřin poté se spustí i nahrávka s rušením. Celkem tak vzniklo 12 záznamů pro testování.



Obrázek 3.9: Schema místnosti, kde byla nahrávána reálná data.

Kapitola 4

Vývoj metody

Zadáním této práce je odstranění signálu z nahrávky, který je známý nebo zjistitelný ze zdroje. Velmi jednoduchým řešením by bylo tento signál ze vstupní nahrávky jednoduše odečíst. Výsledný signál y ze signálu vstupní nahrávky x , kde je známý, respektive ruchový signál z obsažen, by tak byl vypočten pomocí vzorce 4.2.

$$y[k] = x[k] - z[k] \quad (4.1)$$

Tato metoda ovšem může být účinná pouze v případě, že je známý signál ve stejné podobě, jako je obsažen ve vstupním signálu. Při reálných situacích s tím ovšem nemůžeme počítat. Známy signál na vstupní nahrávce je ovlivněný řadou faktorů, které vyplývají z prostředí, kterými se šířil, nebo i vlastnostmi reproduktorů a nahrávacích zařízení. V kapitole 2, především v sekcích 2.1, byly tyto faktory rozebrány.

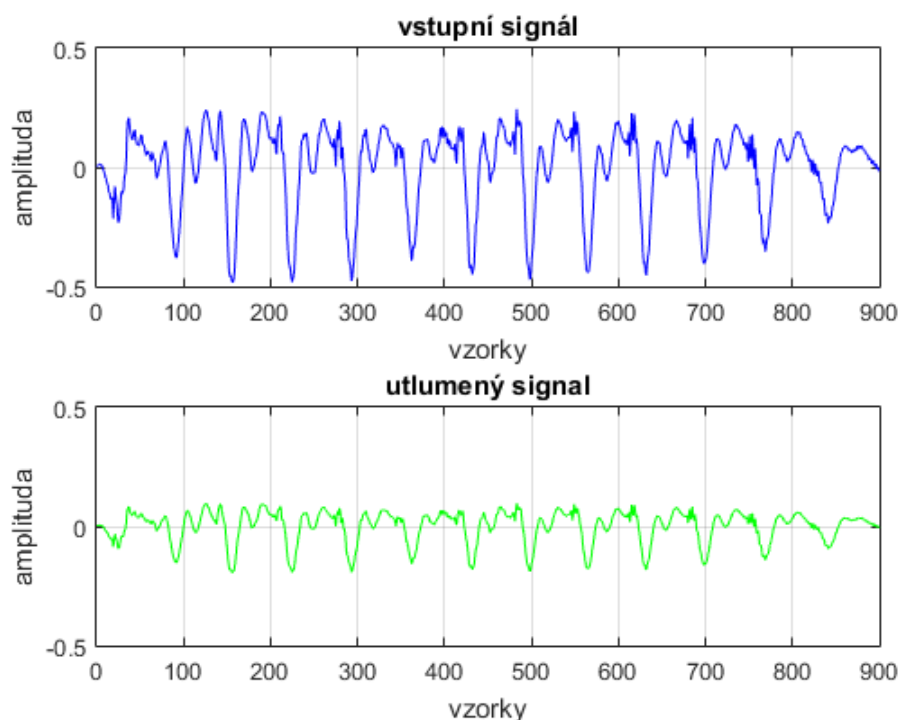
V této části práce si postupně jednotlivé faktory rozebereme a vyzkoušíme a návrhy, jak vzniklé problémy řešit, případně proč jsou řešení nedostačující. Začneme nejprve s jednoduchými vlivy, které jdou v signálu snadno vyhledat a odstranit a budeme postupovat ke složitějším. Kapitola tak může sloužit i jako návod pro řešení obdobných problematik.

4.1 Útlum zvuku v prostředí

Nejvýraznějším faktorem ovlivňující zvuk v prostředí je jeho útlum, tedy snižování intenzity zvuku se zvyšující se vzdáleností od zdroje. V našem případě počítáme, že vzdálenost zdroje zvuku od zařízení, které ho zaznamenává, bude konstantní a můžeme tak považovat za konstantní i útlum. Modelujeme si ho tedy konstantou g (z anglického slova *gain*). Na grafech 4.1 je zobrazen krátký zvukový signál ovlivněný útlumem prostředí. Je patrné, že se útlum projeví jako snížení amplitudy. Toto snížení jde v ideálním případě vypočítat vydělením obou signálů, případně kteréhokoliv vzorku:

$$g = \frac{x[t]}{u[t]} \quad (4.2)$$

kde g je hledaný útlum, $x[t]$ libovolný vzorek vstupního signálu a $u[t]$ odpovídající vzorek utlumeného signálu. To ovšem bude přesné pouze v simulovaném případě. Musíme počítat minimálně se rušením z prostředí. Dalším možným řešením je vypočítat g z podílu energie obou signálů. Tedy podle vzorce



Obrázek 4.1: Útlum signálu při průchodu prostředím.

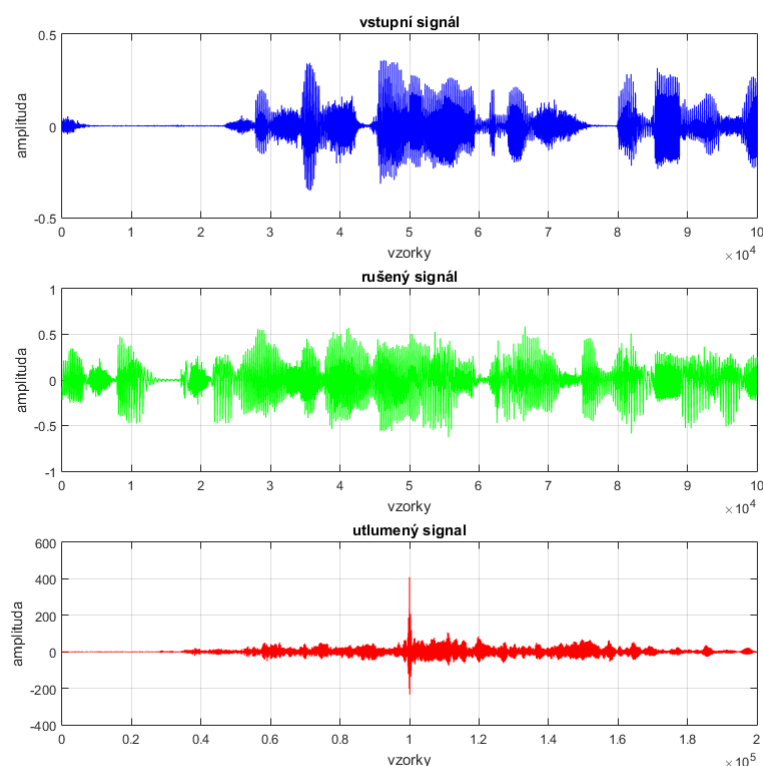
$$g = \sqrt{\frac{E_x}{E_y}}, \quad (4.3)$$

kde E_x je energie signálu x a E_y je energie signálu y . Energie jsou vypočítané podle vzorce 2.5.

Počítat pouze útlum signálu ovšem nemusí být dostačující. Tato práce se zaměřuje především na uzavřené prostory, kde dochází reverberaci a dalším už dříve zmíněným jevům které by při odečítání signálu nevedly k jeho vyčištění, ale spíše ještě dalšímu zarušení. Proto musíme sáhnout k dalším metodám.

4.2 Synchronizace signálů

Synchronizací je v tomto kontextu myšleno nalezení posunutí mezi vstupním signálem a známým signálem. Předpokládejme, že vyhledávaný signál je o proti tomu prohledávanému zpožděný nebo urchylený o časový úsek určený počtem vzorků signálu. Tento počet vzorků potřebuje zjistit. V kapitole 2 byla popsána metoda křížová korelace, která se používá pro zjištění podobnosti dvou signálů. Po provedení této operace získáme křížové korelační koeficienty. Podle těchto koeficientů můžeme zjistit posunutí, tedy lag. Ten je určen nejvyšším křížovým korelačním koeficientem. Problémem v jeho nalezení může být signál, který je rušený. Zvláště v takovém případě, kdy energie rušení výrazně přesahuje energii užitečného signálu. Na obrázku 4.2 jsou zobrazeny dva signály a jejich vzájemná korelace. První signál obsahuje pouze užitečný zvuk, druhý signál obsahuje zvuk, který je výrazně narušený jiným. Signály jsou synchronizované a nejvyšší křížový korelační koeficient by měl být v polovině



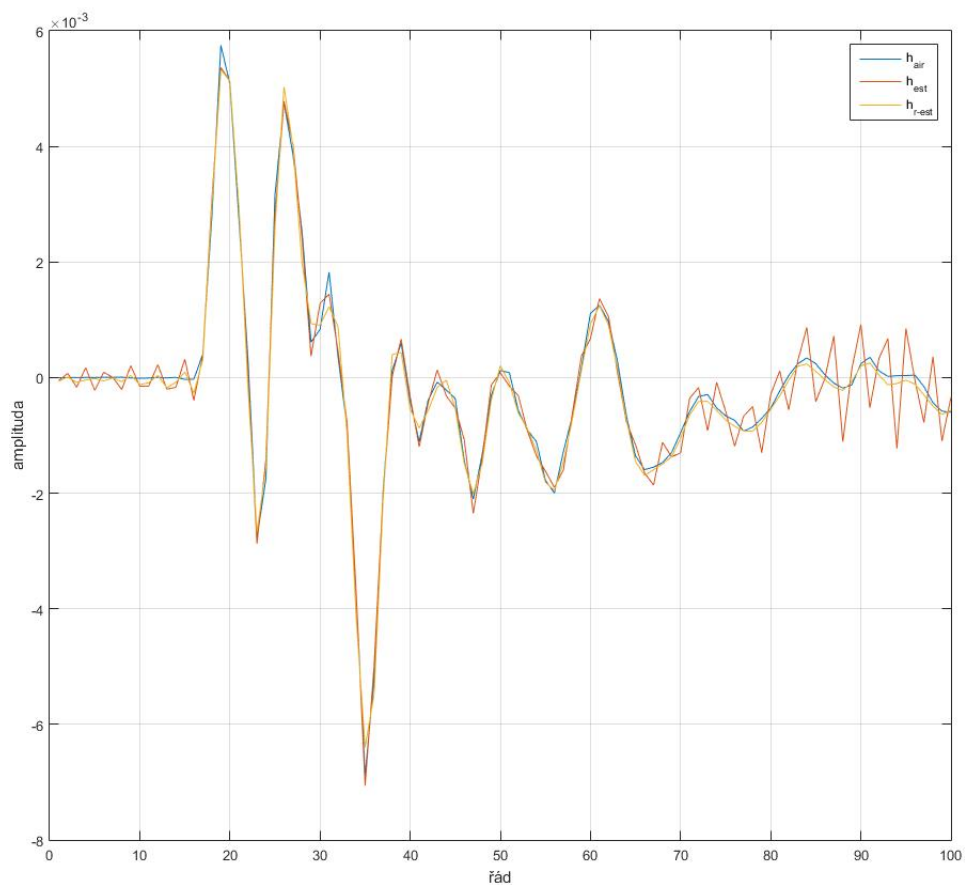
Obrázek 4.2: Vzájemná korelace dvou signálů. Na prvním grafu je signál bez rušení, druhý graf je výrazně rušený jiným signálem. Třetí graf zobrazuje jejich křížovou korelaci.

délky vektoru korelačních koeficientů. Ze třetího grafu je patrné, že tomu tak skutečně je. Z toho vyplývá, že je nalezení lagu úspěšné i v případě, že jsou data významně zarušená. Vzorky signálu musí být ovšem dostatečně dlouhé.

4.3 Výběr částí signálu pro výpočet impulsní odezvy

Původní myšlenkou bylo provádět odhad systému pouze na úsecích signálu, který neobsahuje žádný užitečný signál. Respektive obsahuje pouze rušivý signál ovlivněný místností. Tato technika se ovšem setkala s řadou komplikací. Vyhledání takových úseků bylo poměrně komplikované. Vyhodnocení, které úseky obsahují skutečně pouze rušivý signál, trvalo příliš dlouho a zabíralo většinu délky trvání programu. Navíc i při úspěšném detekování těchto částí signálu nevedl odhad k lepším výsledkům. Impulsní odezvy místností navíc mohou být velmi dlouhé, jak bylo naznačeno v kapitole 2 a vybrané úseky by obsahovaly i vliv částí signálu, o kterých by při výpočtu nebyla žádná informace.

Obrázek 4.3 ukazuje, že na rušeném signálu může být odhad impulsní odezvy lepší, než na čistých datech. Graf na obrázku znázorňuje srovnání prvních sto hodnot třech impulsních odezev. Modře je vyznačena impulsní odezva, která byla získána z [7]. Touto odezvou byl následně filtrována připravená promluva. Nejprve v čisté podobě, poté rušená jiným signálem. Z těchto signálů byla následně odezva vypočtena pomocí Matlab System Identi-



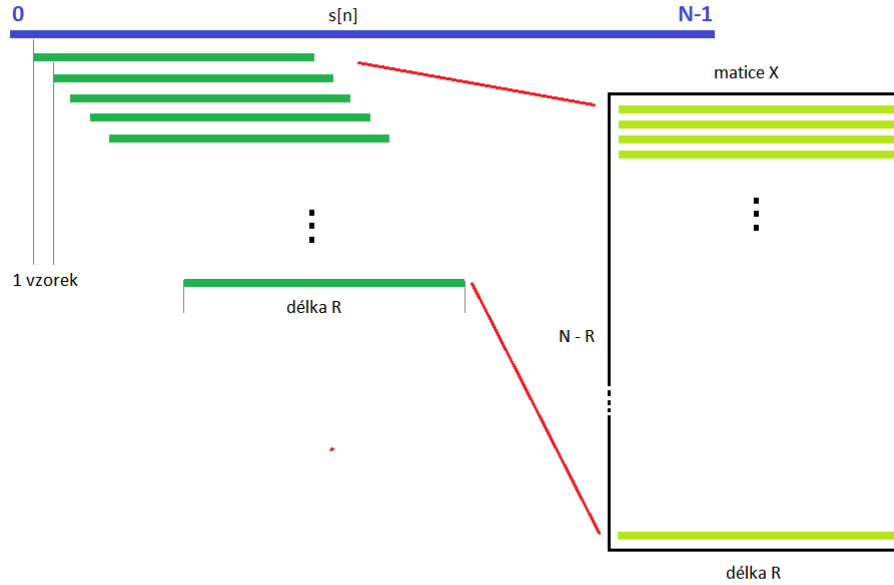
Obrázek 4.3: Srovnání naměřené impulsní odezvy místnosti. h_{air} je původní odezva místnosti, h_{est} odezva místnosti na nerušených datech, h_{r-est} odezva místnosti na rušených datech.

fication Toolbox (dále MSIT). Z výsledku je přitom patrné, že impulsní odezva vypočtená z rušených dat aproximuje původní impulsní odezvu lépe, než ta na čistých datech.

4.4 Výpočet impulsní odezvy

Co to je impulsní odezva, jak vzniká a ovlivňuje zvukový signál, a jak jde z rozdílu dvou signálů zjistit, už je popsáno v kapitole 2 a lze snadno spočítat například MSIT. Tento jednoduchý způsob se ovšem ukázal jako nedostačující pro komplexní zpracování signálů a je potřeba impulsní odezvu vypočítat vlastní metodou. Jak takový výpočet probíhá si nejprve ukážeme v časové doméně. Řekněme, že máme signál x , který obsahuje N vzorků. Pro tento signál si vytvoříme matici X , která bude mít rozměry $R \times (N - R)$, kde R je řád impulsní odezvy, kterou chceme vypočítat. Matici postupně naplňujeme postupem naznačeným na schématu 4.4

Matrice se tedy plní částmi signálu o délce R od počátku signálu. Každá další část je od té předchozí posunuta vždy o jeden vzorek. Ve vektoru je tak vždy současně $x[n]$ a



Obrázek 4.4: Postup naplnění matice pro výpočet impulsní odezvy v časové oblasti.

$R-1$ předchozích vzorků. Pro prvních R vektorů předchozí vzorky neznáme, proto těchto vektorů v matici bude celkem $N - R$. Na tuto vzdálenost upravíme i signál y . Nyní hledáme aproximaci řešení soustavy rovnic zadané matici X , kde je y výsledkem a hledaná impulsní odezva systému h neznámá.

Odhad systému v časové doméně vyhoví, pokud jsou data dokonale synchronizovaná. Ovšem jakékoliv posunutí či nepřesnost v časování signálů nepovede k úspěchu. Proto je třeba provádět odhad systému ve frekvenční oblasti. To má i tu výhodu, že je výpočet výrazně rychlejší, jak už bylo uvedeno v 2.5.

Provádění výpočtu systému ve frekvenční doméně je zdánlivě jednodušší, jelikož impulsní odezva odpovídá pouhému dělení obou signálů. My ovšem pracujeme s dlouhými signály a pro analyzování dělíme na krátké rámce o délce kolem dvaceti milisekund. Impulsní odezva je však kvůli reverberaci mnohem delší a zasahuje tak i do dalších rámců. A naopak, pokud bychom odhadovali systém pouze na takto krátkých úsecích, byl by značně nepřesný, protože by výpočtu chyběla informace o signálu z předchozích rámců. Proto musíme počítat odezvu na základě více rámců zpětně. Jedním z možných řešení je zvolit kompromisní délku rámců. Lepším řešením je ovšem zvolit podobný postup, jako v případě odhadu v časové doméně. Tato technika byla vytvořena na základě rady ing. Kateřiny Žmolíkové [12]. Formálně zapsáno:

$$Y_t[k] = X_t[k] \times H_0[k] + X_{t-1}[k] \times H_1[k] + \dots + X_{t-N}[k] \times H_N[k], \quad (4.4)$$

kde X a Y jsou vektory našich signálů, t odpovídá času v rámcích, N je délka signálu a H značí hledanou impulsní odezvu. Opět tedy hledáme nejlepší řešení rovnic. Tentokrát ovšem nebudeme pracovat s hodnotami signálu v čase, ale hodnotami frekvencí, respektive spektrem jednotlivých rámců. Nejprve si oba signály x a y rozdělíme na krátké rámce o

délce L . Na těchto rámcích provedeme spektrální analýzu pomocí Fourierovy transformace. Výsledky postupně ukládáme do matic. Stačí nám ukládat první polovinu spektra, matice tak budou mít délku $L/2 + 1$. Jejich šířka bude odpovídat počtu rámců R . Jako další krok si sestavíme levou a pravou stranu pro řešení soustavy rovnic. To je provedeno obdobně, jako u výpočtu v časové doméně. Výsledkem je matice impusních odezev, pomocí které je filtrován signál. Buď je možné sestavit matici do jednoho signálu, nebo filtrovat signál po rámcích. Impulsní odezva systému v sobě samozřejmě zahrnuje i útlum popsany v sekci 4.1, takže není nutné útlum modelovat zvlášť. Ovšem vzhledem k časové náročnosti především u výpočtů v časové doméně tím můžeme snížit časovou náročnost.

Kapitola 5

Návrh metody

Na základně zjištění z předchozí kapitoly 4 navrhne konkrétní metodu odstranění šumu z nahrávky. Metoda je popsána algoritmem, který na vstupu očekává vstupní nahrávku, ve které je třeba potlačit známý signál, a nahrávku obsahující pouze ruchový signál. Na výstupu vytvoří nahrávku, kde je šum odstraněn. Algoritmus je dekomponován na jednotlivé části, které řeší konkrétní problematiky.

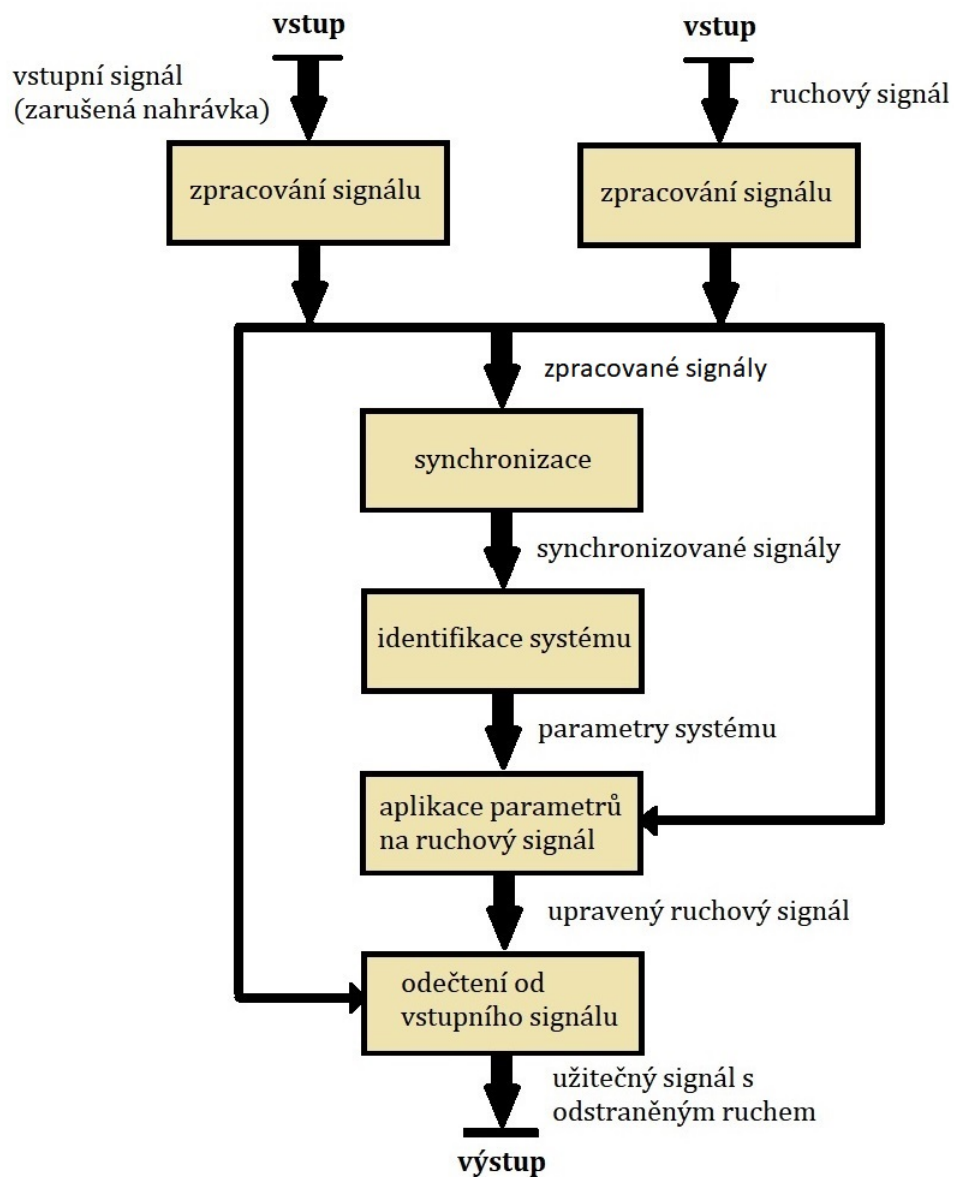
Na obrázku 5.1 popisuje celý navržený postup. Oba signály jsou zpracované a je z nich odstraněna stejnosměrná složka. Takto rozdělené signály jsou pak synchronizovány tak, že je ve vstupním signálu vyhledán začátek a konec ruchového signálu. Jsou vytvořeny dva nové signály o stejné délce. V prvním je pouze ta část vstupní nahrávky, kde je obsažena známý rušivý signál, a v druhém je obsažen pouze onen rušivý signál ze známé nahrávky. Z těchto dvou signálů je dále vypočtena parametry, kterými byly signály ovlivněny. Jakmile jsou parametry zkrácení známy, stačí známý ruchový signál upravit a odečíst od signálu vstupního.

5.1 Synchronizace

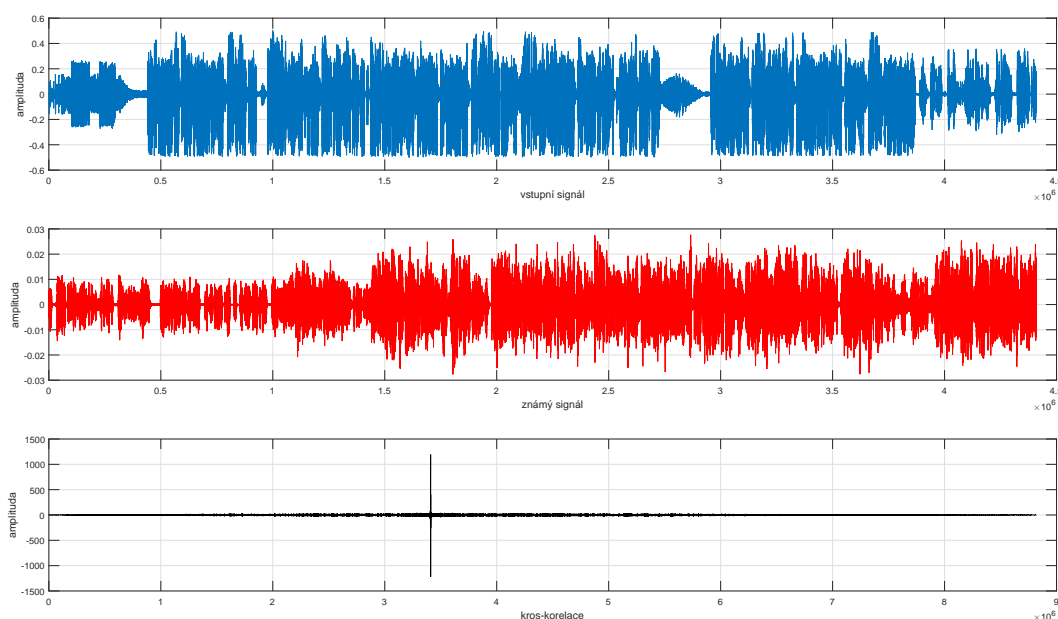
Ze sekce 4.2 vyplývá, že není nutné nijak upravovat vstupní signály před provedením jejich křížové korelace a následné synchronizace. Oba signály musí být ovšem dostatečně dlouhé. Proto pro synchronizaci bude využito celé délky signálů. Křížové korelační koeficienty spolu s příslušnými zvukovými signály je ukázána na obrázku 5.2, kde je na horním grafu zobrazena nahrávka řeči, rušená hudbou na prostředním grafu. Hudba přitom začíná až několik sekund po začátku řeči. Na posledním grafu jsou zobrazeny křížové korelační koeficienty. Jeden tento koeficient má výrazně vyšší amplitudu než ostatní a určuje tak hledané zpoždění. Na základě něj je vypočítán začátek a konec rušivého signálu ve vstupní nahrávce.

5.2 Identifikace systému

Identifikace systému proběhne ve frekvenční oblasti. Oba synchronizované signály jsou rozděleny na rámce, které jsou pomocí Fourierovy transformace převedeny do jejich spekter. Délka rámců je nastavena tak, aby odpovídala dvaceti milisekundám záznamu s deseti milisekundovým překrytím. Z vypočtených spekter jsou sestaveny dvě matice. Z obou matic je odstraněna imaginární složka. Z matice se vstupním signálem je následně sestavena levá strana výpočtu impulsní odezvy, z matice s ruchovým signálem jsou vybírány vektory pro



Obrázek 5.1: Schéma postupu navržené metody.



Obrázek 5.2: Korelace dvou na vzájem posunutých zvukových signálů.

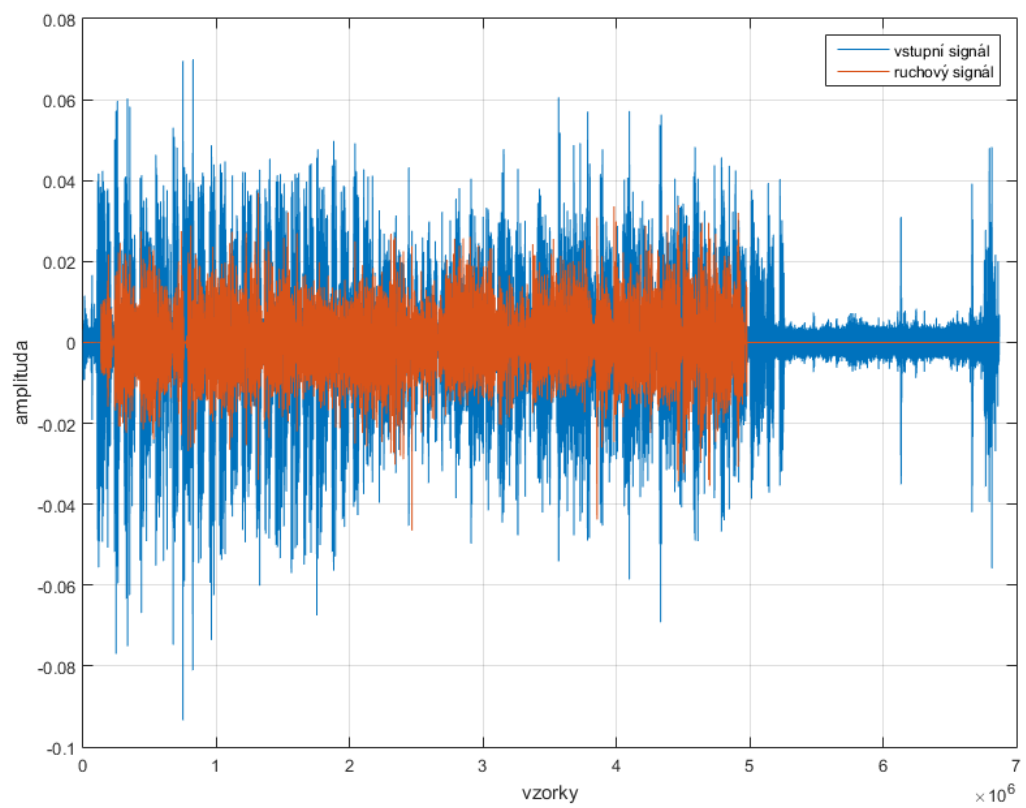
pravou stranu výpočtu. Výpočet pak probíhá podle popisu v sekci . Výsledkem je matice obsahující frekvenční charakteristiky signálu, který můžeme aplikovat na signál.

5.3 Aplikace na ruchový signál

Aplikace na signál opět probíhá po rámcích. Ruchový signál je rozdělen do stejně velkých rámců, jako při identifikaci systému. Dále je rovněž provedena Fourierova transformace a odstraněna imaginární část, která je uložena pro syntézu. Prochází se jednotlivými rámci, které jsou filtrovány připravenými charakteristikami. Výsledek musí být znovu syntetizován do celého signálu. Nejprve je navracena uložena imaginární část, poté je provedena zpětná Fourierova transformace. Výsledek je sestaven do jednoho vektoru a vytvořen tak ruchový signál o stejné podobě, jako je na nahrávce. Na obrázku 5.3 je zobrazena ukázka vstupního signálu a upraveného ruchového signálu připraveného k odečtení.

5.4 Odečtení od vstupního signálu

Odečtení ruchového signálu od vstupní nahrávky probíhá opět ve frekvenční oblasti. Znovu jsou oba signály rozděleny na rámce o délce dvacet milisekund, tentokrát ovšem bez překrytí. Na rámcích je provedena Fourierova transformace a odstraněna imaginární složka. Od spekter rámců vstupního signálu jsou odečtena spektra rámců upraveného ruchového signálu. Musíme si dát pozor, aby výsledek nebyl menší než nula. Pokud se tak stane, dosadíme místo záporné hodnoty nulu. K výsledku je zpět doplněna imaginární složka, provedena inverzní Fourierova transformace a z matice je znovu složen jeden signál. Tento signál je náš požadovaný výsledek, který by neměl obsahovat žádný obsah ruchového signálu.



Obrázek 5.3: Zobrazení vstupního signálu (modře) a upraveného ruchového signálu (červeně).

Kapitola 6

Implementace

Pro implementaci programu byl zvolen MATLAB. MATLAB je vysokoúrovňová platforma určená pro řešení inženýrských a vědeckých problémů [11]. Hlavní důvody pro zvolení tohoto programovacího jazyka je propracované vývojové prostředí, zejména vestavěné funkce pro práci se signály a grafy. Také má rozsáhlou knihovnu funkcí potřebných pro řešení obdobných problémů, kvalitní dokumentaci a širokou komunitu uživatelů. Program je implementován jako knihovní funkce, která očekává na vstupu tři parametry:

- vstupní signál - signál zvukové nahrávky rušené známým signálem, který potřebujeme odstranit
- známý signál - signál nahrávky rušení obsaženého ve vstupním signálu
- vzorkovací frekvence - vzorkovací frekvence obou signálů musí být stejná

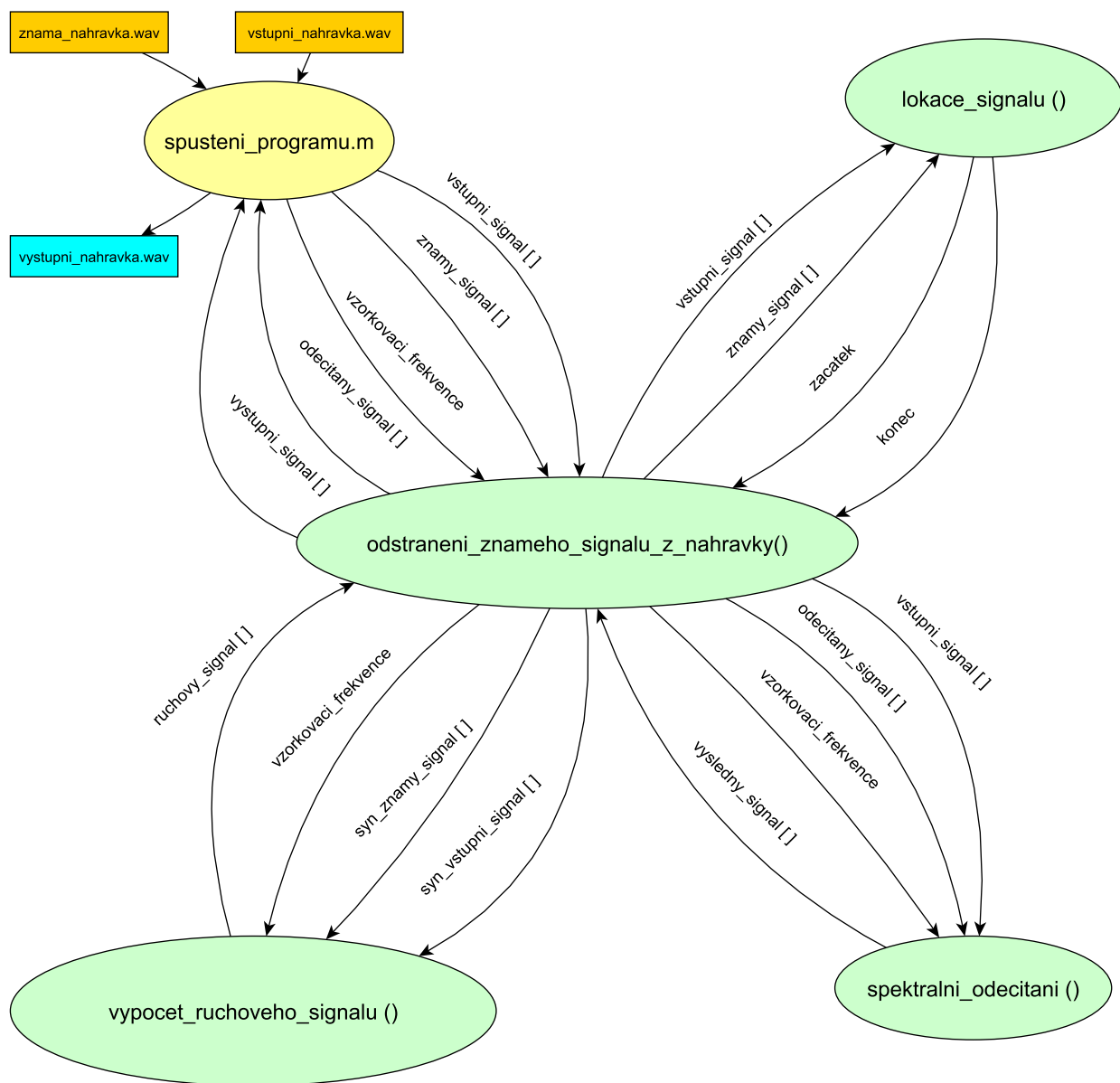
Oba zvukové soubory musí být pro správnou funkčnost programu vzorkovány o stejné vzorkovací frekvenci. Parametry vstupní signál a známý signál jsou očekávány ve sloupcovém vektoru tak, aby mohl být přímo použit výstup z vestavěné funkce *audiowrite()*. Na výstupem funkce jsou dva řádkové vektory:

- výstupní signál - výsledný signál, ze kterého je odstraněno rušení známým signálem
- odečítaný signál - signál obsahující odstraněné rušení

Tyto vektory obsahují záznam zvuku o stejné vzorkovací frekvenci, jako byly vstupní signály. Použití programu je ukázáno v souboru `spusteni_programu.m`, který je popsán v sekci 6.5. Program je strukturován do funkcí, které řeší jednotlivé problematiky popsané v kapitole 4. Struktura je modulární, tedy řešena tak, aby na sobě jednotlivé fáze byly nezávislé a mohly být případně nahrazovány efektivnějšími metodami. To však vede k tomu, že mohou být některé fáze zpracování signálu opakovány v každém modulu zvlášť. Kompletní struktura programu včetně vstupů a výstupů jednotlivých modulů je zobrazena na obrázku 6.1.

6.1 Odstranění známého signálu z nahrávky

Jak napovídá název, jde o klíčovou funkci celého programu. Na vstupu má dvě pole obsahující vstupní signál a známý signál. Posledním parametrem je vzorkovací frekvence. Na



Obrázek 6.1: Struktura programu.

začátku je volána funkce lokace signálu, která má za parametry vstupní a známý signál. Tato funkce vrátí pozici známého signálu ve vstupní nahrávce ve dvou proměnných začatek a konec jako indexy do pole obsahující vstupní signál. Na základě těchto hodnot jsou vytvořena dvě nová pole o stejné velikosti, která obsahují části vstupního a známého signálu tak, aby obsahovala stejné části signálu. Respektive, aby celá část známého signálu ve vstupním signálu byla obsažena v novém poli a aby pole neobsahovala signál, který známý signál neobsahuje. Tato pole jsou společně se vzorkovací frekvencí vstupními parametry funkce *vypocet_ruchoveho_signalu()* 6.3. Tato funkce vrací upravený známý signál do podoby, v jaké je obsažen ve vstupním signálu. Poté už stačí tento signál od vstupního signálu odečíst na místě určeném hodnotami začatek a konec. Odečítání se musí provést ve spektrální oblasti, k tomu je vytvořena funkce *spektralni_odecitani()* popsaná níže 6.4.

6.2 Lokace signálu

Funkce sloužící pro nalezení známého signálu ve vstupních datech. Prohledávaným signálem je v našem případě vstupní nahrávka a hledaným signálem známý signál. Základem funkce je operace kross-korelace, která oba signály porovnává. Je využita vestavěná funkce MATLABu *xcorr()*, které jsou předány oba signály jako parametry. Operace *xcorr()* vrátí korelační koeficienty a lag, tedy zpoždění hledaného signálu oproti prohledávanému. Z tohoto lagu je potřeba vypočítat indexy do pole prohledávaného signálu. K tomu se využije funkce *max()*, která nalezne největší korelační koeficient a jeho pozici, která nás zajímá. Tento koeficient může být i záporný, proto je jsou korelační koeficienty předávány funkci *max()* v absolutních hodnotách, které jsou získány pomocí funkce *abs()*. Index začátku tak odpovídá pozici nejvyššího korelačního koeficientu. Index konce je pak součet indexu začátku s délkou hledaného signálu. Pokud je hledaný signál delší než prohledávaný signál, je jako index konce vrácena délka prohledávaného signálu. Protože impulsní odezvu počítáme ve frekvenční oblasti, potřebujeme provést spektrální analýzu fourierovou transformací. Navíc chceme počítat impulsní odezvu z více rámců zpět, proto si musíme ukládat spektra rámců do matice. Alokujeme si tedy dvě matice, pro každý signál jednu, kde každý sloupec bude reprezentovat polovinu spektra každého rámce.

6.3 Výpočet ruchového signálu

Toto je klíčová funkce celého programu. Ze vstupního a známého signálu vypočítá impulsní odezvu systému místnosti popsané v kapitole 2. Pro zjištění popsané v kapitole 4 bude odhad systému probíhat ve frekvenční oblasti. Také bude probíhat na celém signálu, nikoliv pouze na rámcích bez rušení, jako bylo původně zamýšleno. Z obou signálů je nejprve odečtena stejnosměrná složka. To je provedeno odečtením průměrné hodnoty signálu od celého příslušného vektoru. Ke zjištění průměrné hodnoty je využita funkce *mean()*. Dále jsou oba signály rozčleněny na rámce. Délka rámců byla sice zvolena na dvacet milisekund, ovšem experimenty s délkou rámců ukázaly, že jejich menší změny nemají na odhad systému významný vliv. Naopak implementace Fourierovy transformace pracuje nejrychleji na signálech o délce odpovídající mocnině čísla dvou [1], proto je délka rámců upravena na nejbližší tuto mocninu zjištěnou funkcí *nextpow2()*. Pro rychlejší práci algoritmu je také lepší si nejprve připravit a alokovat struktury pro zápis jednotlivých výsledků a mezivýsledků. Hlavním principem činnosti této funkce je sestavení matic pro řešení soustavy rovnic, kde jako neznámá vystupuje vektor hledané impulsní odezvy. Vytváření těchto matic je poměrně

nepřehledné, proto si je ukážeme na samotném kódu. Nejprve vytvoříme matice obsahující spektrum obou signálů do proměnných *mat_X* a *mat_Y*:

```
for ii = 1:pocet_ramcu,

    % aktualni ramec
    ramec_X = vstupni_signal(od:do);
    ramec_Y = znamy_signal(od:do);

    % potlacení okraje pomocí Hannova okna
    hann_ramec = hann(Ls);
    ramec_X = ramec_X .* hann_ramec';
    ramec_Y = ramec_Y .* hann_ramec';

    % fourierova transformace X
    fft_ramec_X = fft(ramec_X);
    fft_ramec_Y = fft(ramec_Y);

    mat_X(:,ii) = fft_ramec_X(1:Ls/2+1)';
    mat_Y(:,ii) = fft_ramec_Y(1:Ls/2+1)';

    % nový rozsah ramcu
    od = od + posun;
    do = do + posun;

end
```

Základem je cyklus *for*, který prochází jednotlivé rámce matice. Rámce jsou vytvořeny pomocí hannova okna, které utlumí signál na okrajích rámců. Pro rámce obou signálů je provedena fourierova transformace pomocí funkce *fft*. Tato funkce vrátí spektrální analýzu rámců o stejné délce, jaká je délka původních rámců. Spektrum je zrcadlené, proto nám stačí pracovat s pouze jednou polovinou. Do proměnných *mat_X* a *mat_Y* se tak ulouží pouze vzorky od jedné do poloviny *Ls*, která obsahuje hodnotu počtu vzorků celého rámce. Dalším krokem je vypočtení hledané impulsní odezvy řešením soustavy rovnic. K tomu je použita funkce *mldivide()*, která je v kódu zkráceně volána pomocí znaku `\` v následujícím formátu:

$$A \times x = B \quad (6.1)$$

kde *x* je hledaná proměnná, *A* je levá strana soustavy rovnic a *B* pravá. V našem případě budeme tuto funkci využívat následujícím způsobem:

$$\text{mat_LEVA} \times h_est = prava \quad (6.2)$$

kde *h_est* je hledaná impulsní odezva ve frekvenční oblasti, proměnné *mat_LEVA* a *prava* jsou naplněny následne:


```

for kk = 1:Ls/2+1,

    % postaveni matice levych stran
    ii = 1;
    for n = h_pocet:pocet_ramcu
        mat_LEVA(ii,:) = abs_mat_Y(kk,n:-1:(n-h_pocet+1));
        ii = ii + 1;
    end

    % tvorba prave strany - jen dlouhy sloupcovy vektor
    prava = abs_mat_X(kk, h_pocet:pocet_ramcu)';

    % odhad impulsni odezvy
    h_est = mat_LEVA \ prava;

    mat_impulsni_odezvy(kk,:) = h_est';

end

```

Matice *mat_LEVA* je tak vytvořena z matice *abs_mat_Y*, což je matice *mat_Y* převedena do absolutní hodnoty. Příkaz

```
abs_mat_Y(kk,n:-1:(n-h_pocet+1))
```

vrátí několik hodnot daných proměnnou *h_pocet* zpětně z předchozích rámců o stejné frekvenci. Takto je postupně naplněna celá matice. Vektor *prava* obsahuje v každém cyklu vždy shodné frekvence z analýzy celého signálu vybrané z matice *abs_mat_X*. Řešením rovnice vznikají úseky impulsní odezvy *h_est*, kterými je postupně naplňována předem alokovaná matice *mat_impulsni_odezvy*. Máme tak zjištěnou impulsní odezvu z více rámců. Nyní je potřeba filtrovat známý signál touto odezvou. Opět budeme procházet po jednotlivých frekvencích spektra a filtrovat samostatně:

```

for kk = 1:Ls/2+1,
    mat_ruchovy_signal(kk,:) = filter(mat_impulsni_odezvy(kk,:)',
        1, abs_mat_Y(kk,:));
end

```

Nyní máme v matici *mat_ruchovy_signal* obsažený filtrované spektrum signálu. Nyní ho můžeme sestavit třeba ho znovu sestavit. Jelikož jsme pracovali s absolutními hodnotami signálu, přišli jsme o informaci o fázi, kterou je třeba do signálu znovud doplnit. Nyní můžeme provést syntézu získaného signálu, tedy pomocí zpětné Fourierovy transformace (funkce *ifft()*). Postupujeme po rámcích a rámce si doplníme o druhou polovinu zrcadlením té první. Funkce funkce *ifft()* vrátí kvůli nepřesnostem ve výpočtu znovu komplexní čísla, imaginární hodnoty jsou ovšem velmi malé. Odstraníme je příkazem *real()*. Výsledkem je ruchový signál, který potřebujeme z původní nahrávky vyčistit.

6.4 Odečítání ve spektru

Jak bylo ukázáno v kapitole 4, signály je třeba odečítat ve spektrální oblasti. Pro tento účel byla vytvořena funkce *spektralni_odecitani()*, která má na vstupu dva vstupní signály x a y , a vrací jejich rozdíl $x-y$. Dále je funkci třeba zadat vzorkovací frekvenci, která musí být u obou signálů stejná. Funkce rozdělí signál na rámce o přibližné délce odpovídající přibližně dvaceti milisekundám ve zvukovém záznamu bez překrytí. Na každém rámci obou signálů je provedena fourierova transformace. Výsledek je rozdělen na část modulu pomocí funkce *abs()* a na část argumentu pomocí funkce *angle()*. Modul signálu y je následně odečten od modulu signálu x . Je třeba si pohlídat, aby výsledná hodnota nebyla menší než nula. Při záporných hodnotách je proto místo výsledku dosazena nulová hodnota. Do výsledného rámce je zpět doplněn původní argument x a rámce jsou zpět sestaveny do celého signálu.

6.5 Spuštění programu

Pro ukázkou spuštění funkce byl vytvořen soubor *spusteni_programu.m*. Tento soubor slouží jako prezentace použití programu, práce s jeho rozhraním, načítání dat ve správném formátu či základní testování výstupu programu. Funkce načítá vstupní data a upravuje do formátu vyžadovaného funkcí. Data jsou načítána ve zvukovém formátu *wav*. Vstupem jsou dva soubory, odposlechnutou nahrávku, respektive vstupní signál, a signál obsahující známou nahrávku. Funkce oba soubory načte pomocí funkce *audioread()*, která je součástí MATLABu, a uloží je do proměnných *vstupni_nahravka* a *znama_nahravka*. Tato pole jsou sloupcové vektory typu *double*. Dále z obou zvukových záznamů zjistí vzorkovací frekvenci jako jedno číslo typu *double*. Tyto frekvence porovná a pokud nejsou shodné, vypíše chybu a ukončí program. V MATLABu je ovšem možné signál podvzorkovat pomocí funkce *decimate()*. Pokud jsou všechny parametry ve správném formátu, je možné spustit funkci *odstraneni_znameho_signalu_z_nahravky()*. Ta vrací výsledný signál v řádkovém vektoru, který se zapíše do souboru *vystupni_nahravka.wav* pomocí funkce *audiowrite()* se stejnou vzorkovací frekvencí jako vstupní nahrávka. Dále je zde ukázáno volání funkcí pro testování výstupu.

Kapitola 7

Vyhodnocení a testování

Jak bylo řečeno v kapitole 3, vyhodnocení bude probíhat na dvou sériích dat. Tedy nad simulovanými daty a reálnými daty. Pro vyhodnocení simulovaných dat budou použity techniky SNR a kepstrální vzdálenost popsané v části 2.7. Pro vyhodnocení reálných dat bude použita pouze kepstrální vzdálenost (dále CD). Vyhodnocení CD probíhá vůči originálnímu záznamu. Pro vyhodnocení SNR je originální záznam upraven známou impulsní odezvou příslušné místnosti.

Nejprve si ukážeme, jak si implementovaný program vede v situaci, kdy vstupní signál není známou nahrávkou nijak zarušen. Výsledek by měl být totožný se vstupem, ovšem vzhledem k tomu, jakým způsobem naše metoda pracuje, respektive nějaký signál se odečítá v každém případě, bude výsledek vždy trochu pozměněn. Naměřené hodnoty CD zobrazené v tabulce na obrázku 7.1 jsou však velmi nízké. Po přehrání výsledného záznamu je místy slyšitelné praskání, zvláště u rockové hudby. To je však možné označit za přijatelné a nebrání srozumitelnosti nahrávky. Tyto naměřené hodnoty je tak možno brát jako srovnání a výsledky programu u rušených nahrávek by se v ideálním případě měly těmito hodnotám blížit.

| Bez rušení | |
|----------------|------------------------|
| | CD (*10 ⁹) |
| klasická hudba | 0,7617 |
| rocková hudba | 2,1266 |
| rap | 0,1627 |
| mluvené slovo | 0,0369 |
| průměr | 0,7720 |

Obrázek 7.1: Tabulka naměřených hodnot pro nahrávky, které nejsou známým signálem rušeny.

| aula: | | | | | | | | |
|-------------------------|----------|------------------------|----------|------------------------|----------|------------------------|----------|------------------------|
| faktor zesílení rušení: | 0.2 | | 0.5 | | 1 | | 2 | |
| rušení: | SNR [db] | CD (*10 ⁹) | SNR [db] | CD (*10 ⁹) | SNR [db] | CD (*10 ⁹) | SNR [db] | CD (*10 ⁹) |
| klasická hudba | 43,1789 | 0,1809 | 41,9668 | 0,5399 | 39,7258 | 1,3392 | 36,0328 | 2,9971 |
| rocková hudba | 41,6035 | 4,4570 | 37,4554 | 5,8720 | 32,7803 | 5,7632 | 27,3518 | 4,3888 |
| rap | 38,2366 | 4,1144 | 32,5400 | 9,6525 | 27,2224 | 12,5140 | 21,4946 | 13,8591 |
| mluvené slovo | 42,7820 | 0,4857 | 40,5383 | 0,9311 | 37,1284 | 2,0304 | 32,4581 | 3,7741 |
| průměr | 41,4503 | 2,3095 | 38,1251 | 4,2489 | 34,2142 | 5,4117 | 29,3343 | 6,2548 |

Obrázek 7.2: Tabulka naměřených hodnot simulovaných dat pro místnost aula

| chodba: | | | | | | | | |
|-------------------------|----------|------------------------|----------|------------------------|----------|------------------------|----------|------------------------|
| faktor zesílení rušení: | 0.2 | | 0.5 | | 1 | | 2 | |
| rušení: | SNR [db] | CD (*10 ⁹) | SNR [db] | CD (*10 ⁹) | SNR [db] | CD (*10 ⁹) | SNR [db] | CD (*10 ⁹) |
| klasická hudba | 52,9548 | 0,0067 | 51,3861 | 0,0607 | 49,3876 | 0,2668 | 46,8284 | 0,9197 |
| rocková hudba | 51,4247 | 2,2076 | 48,2884 | 3,9717 | 45,4918 | 8,6824 | 42,1726 | 12,6890 |
| rap | 48,3637 | 0,3226 | 44,8271 | 2,6087 | 41,5890 | 6,2905 | 37,5320 | 9,2096 |
| mluvené slovo | 52,6259 | 0,1036 | 50,4194 | 0,1183 | 48,1165 | 0,2301 | 45,5872 | 1,0303 |
| průměr | 39,2514 | 0,6601 | 48,7303 | 1,6899 | 46,1462 | 3,8675 | 43,0301 | 5,9621 |

Obrázek 7.3: Tabulka naměřených hodnot simulovaných dat pro chodbu

7.1 Vyhodnocení simulovaných dat

Pro testování bylo připraveno 64 různých nahrávek se simulovanými daty, každá vyhodnocena dvěma technikami. Celkem tak bylo získáno 128 hodnot. Nejprve si vyhodnotíme výsledky podle jednotlivých místností. Tabulka 7.2 obsahuje výsledky pro aulu, tabulka 7.3 pro chodbu, tabulka 7.4 pro zasedací místnost a 7.5 pro přednáškovou místnost. Tabulka 7.6 pak zobrazuje průměrné hodnoty vypočtené pro jednotlivé místnosti. Z vypočtených dat vyplývá, že nejlepší výsledky měla metoda v případě, že byla použita impulsní odezva zasedací místnosti. Dobré výsledky byly i v případě chodby. Naopak aula a přednášková místnost měla výsledky horší. Testovaná metoda si tak poradí lépe s menšími místnostmi, než s těmi většími.

Nic překvapivého nepřináší průměrné výsledky podle toho, jak bylo rušení známým signálem silné. Tyto výsledky jsou zobrazené v tabulce 7.7. Čím silnější rušivý signál obsažený ve vstupních datech byl, tím hůře se s ním metoda vyrovnala.

Nejzajímavější je však situace je však ve srovnání hodnot podle typu rušení zobrazené v tabulce 7.8. V kapitole 3 jsou jednotlivé typy rušení seřazeny podle očekávané obtížnosti odstranění. Klasická hudba měla být nejjednodušší, naopak rušení mluveným slovem mělo být z nahrávky odstranitelné nejhůře. Předpoklad se v případě klasické hudby potvrdil. Průměrná hodnota SNR i CD je u tohoto typu rušení skutečně nejlepší. V případě mluveného slova jsou však výsledky velmi podobné a předpoklad, že mluvený projev bude nejvíce

| zasedací místnost: | | | | | | | | |
|-------------------------|----------|------------------------|----------|------------------------|----------|------------------------|----------|------------------------|
| faktor zesílení rušení: | 0.2 | | 0.5 | | 1 | | 2 | |
| rušení: | SNR [db] | CD (*10 ⁹) | SNR [db] | CD (*10 ⁹) | SNR [db] | CD (*10 ⁹) | SNR [db] | CD (*10 ⁹) |
| klasická hudba | 21,9914 | 0,0457 | 20,1246 | 0,1391 | 17,5728 | 0,4824 | 14,0123 | 1,5168 |
| rocková hudba | 20,1122 | 2,3718 | 16,0697 | 3,2600 | 11,7725 | 3,9249 | 6,7413 | 3,8016 |
| rap | 16,0656 | 0,5584 | 10,2542 | 3,9621 | 5,0086 | 8,8312 | -0,6454 | 12,3700 |
| mluvené slovo | 21,7956 | 0,0642 | 19,0811 | 0,0969 | 15,6757 | 0,2825 | 11,2685 | 1,0942 |
| průměr | 19,9912 | 0,7600 | 16,3824 | 1,8645 | 12,5074 | 3,3802 | 7,8442 | 4,6957 |

Obrázek 7.4: Tabulka naměřených hodnot simulovaných dat pro zasedací místnost

| přednášková místnost: | | | | | | | | |
|-------------------------|----------|------------------------|----------|------------------------|----------|------------------------|----------|------------------------|
| faktor zesílení rušení: | 0.2 | | 0.5 | | 1 | | 2 | |
| rušení: | SNR [db] | CD (*10 ⁹) | SNR [db] | CD (*10 ⁹) | SNR [db] | CD (*10 ⁹) | SNR [db] | CD (*10 ⁹) |
| klasická hudba | 18,7928 | 0,1136 | 16,7186 | 0,4288 | 13,7076 | 1,1257 | 9,5119 | 2,6553 |
| rocková hudba | 16,7854 | 2,4420 | 12,1319 | 3,3003 | 7,3120 | 3,8897 | 1,8601 | 3,6641 |
| rap | 11,3468 | 3,8987 | 4,5539 | 9,5061 | -1,1352 | 12,0900 | -7,0184 | 13,2500 |
| mluvené slovo | 18,7056 | 0,1057 | 15,4860 | 0,4621 | 11,3691 | 1,6514 | 6,2860 | 3,7064 |
| průměr | 16,4077 | 1,6400 | 12,2226 | 3,4243 | 7,8134 | 4,6892 | 2,6599 | 5,8190 |

Obrázek 7.5: Tabulka naměřených hodnot simulovaných dat pro zasedací místnost

| místnost: | SNR [db] | CD (*10 ⁹) |
|----------------------|----------|------------------------|
| přednášková místnost | 9,7759 | 3,8931 |
| aula | 35,7810 | 4,5562 |
| zasedací místnost | 14,1813 | 2,6751 |
| chodba | 44,2895 | 3,0449 |

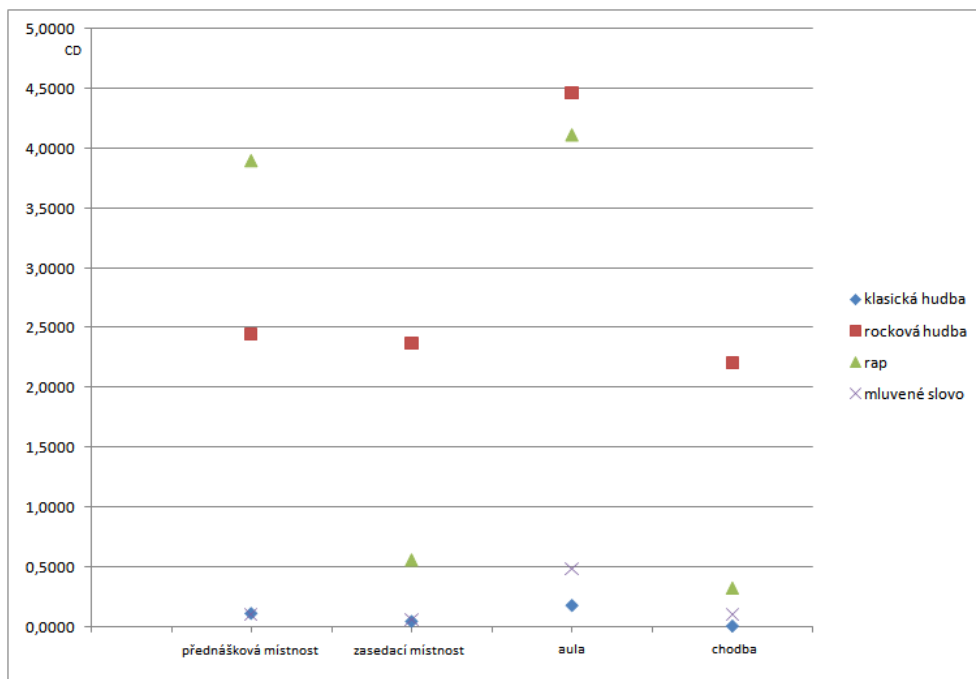
Obrázek 7.6: Průměrný SNR a CD jednotlivých místností.

| faktor | SNR [db] | CD (*10 ⁹) |
|--------|----------|------------------------|
| 0,2 | 29,2751 | 1,3424 |
| 0,5 | 28,8651 | 2,8069 |
| 1 | 25,1703 | 4,3371 |
| 2 | 20,7171 | 5,6829 |

Obrázek 7.7: Průměrný SNR a CD podle koeficientu zesílení rušení.

| rušení: | SNR [db] | CD (*10 ⁹) |
|----------------|----------|------------------------|
| klasická hudba | 30,8683 | 0,8011 |
| rocková hudba | 26,2096 | 4,6679 |
| rap | 20,6397 | 7,6899 |
| mluvené slovo | 29,3327 | 1,0104 |

Obrázek 7.8: Průměrný SNR a CD podle typu rušení.

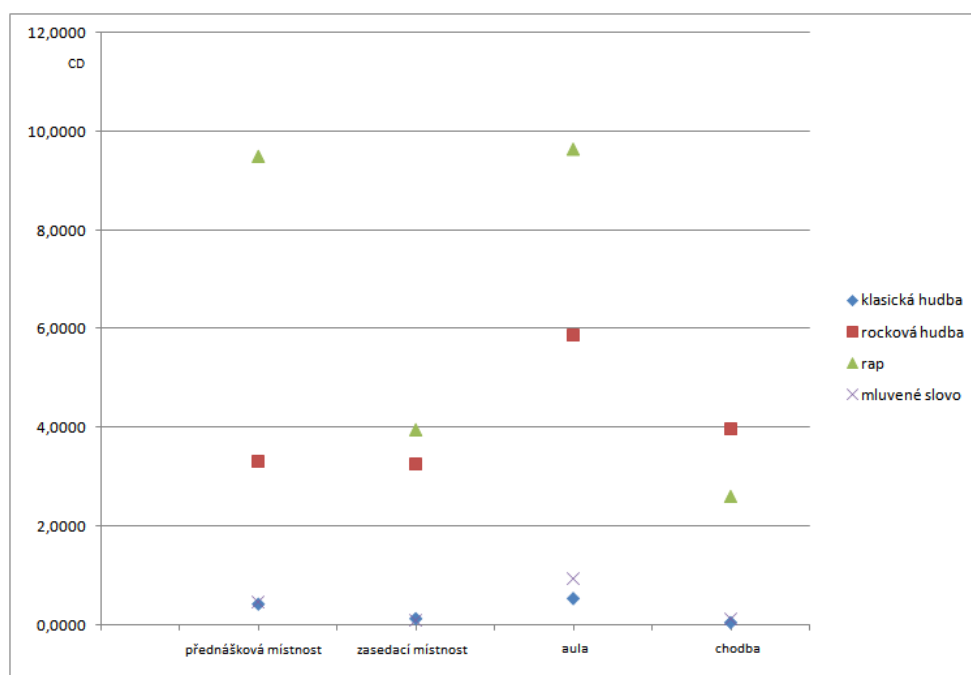


Obrázek 7.9: Graf výsledků CD simulovaných dat pro faktor zesílení rušení 0,2.

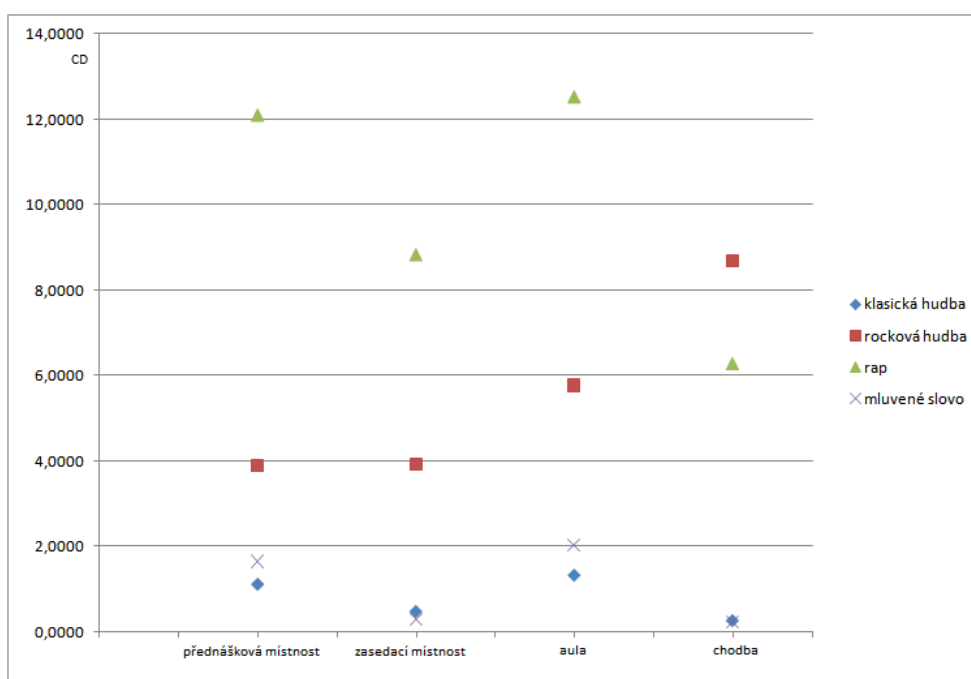
rušit jiný mluvený projev, se neukázal jako správný. Hodnoty CD jsou navíc velmi nízké a v průměru se dokonce blíží hodnotám zcela bez rušení (viz tabulka 7.1) a v některých dílčích výsledcích je dokonce předčila. Výsledky se navíc zdají jednoznačné a konzistentní, takže metoda tak jde pro tento typ rušení vyhodnotit jako velmi úspěšná. To potvrzuje i subjektivní poslech. Výsledné nahrávky obsahují pouze praskání, které bylo patrné i na výsledných nahrávkách bez rušení. Pokud je hudba či druhý mluvčí slyšitelný, tak pouze velmi slabě.

Nejhůře naopak dopadlo rušení rapem. V některých případech bylo SNR při tomto rušení záporné. Z čehož se dá vyvodit, že výsledek po pokusu o odstranění známé nahrávky byl horší, než na nahrávce vstupní. Při poslechu je ovšem patrné, že rapová hudba ruší mluvený projev skutečně velmi silně a na vstupních nahrávkách nebyl užitečný signál mluvčích vůbec slyšitelný. V některých případech si přitom metoda poradila obdobně, jako při klasické hudbě. Je tak možné, že při špatných výsledcích nedošlo vůbec k synchronizaci signálů.

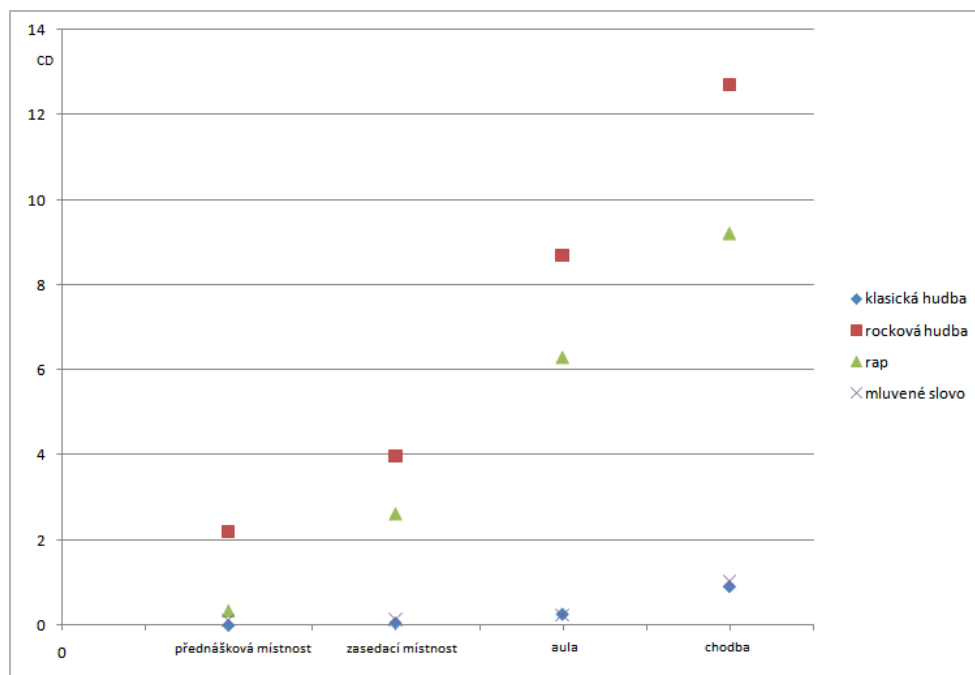
Graf 7.9 zobrazuje závislost zjištěného CD na typu místnosti pro různá rušení a faktor zesílení rušení 0,2. Graf 7.10 zobrazuje závislost zjištěného CD na typu místnosti pro různé druhy typů rušení a faktor zesílení rušení 0,5, graf 7.10 zobrazuje závislost zjištěného CD na



Obrázek 7.10: Graf výsledků CD simulovaných dat pro faktor zesílení rušení 0,5.



Obrázek 7.11: Graf výsledků CD simulovaných dat pro faktor zesílení rušení 1.



Obrázek 7.12: Graf výsledků CD simulovaných dat pro faktor zesílení rušení 2.

typu místnosti pro různé druhy typů rušení a faktor zesílení rušení 1 a graf 7.12 zobrazuje závislost zjištěného CD na typu místnosti pro různé druhy typů rušení a faktor zesílení rušení 2.

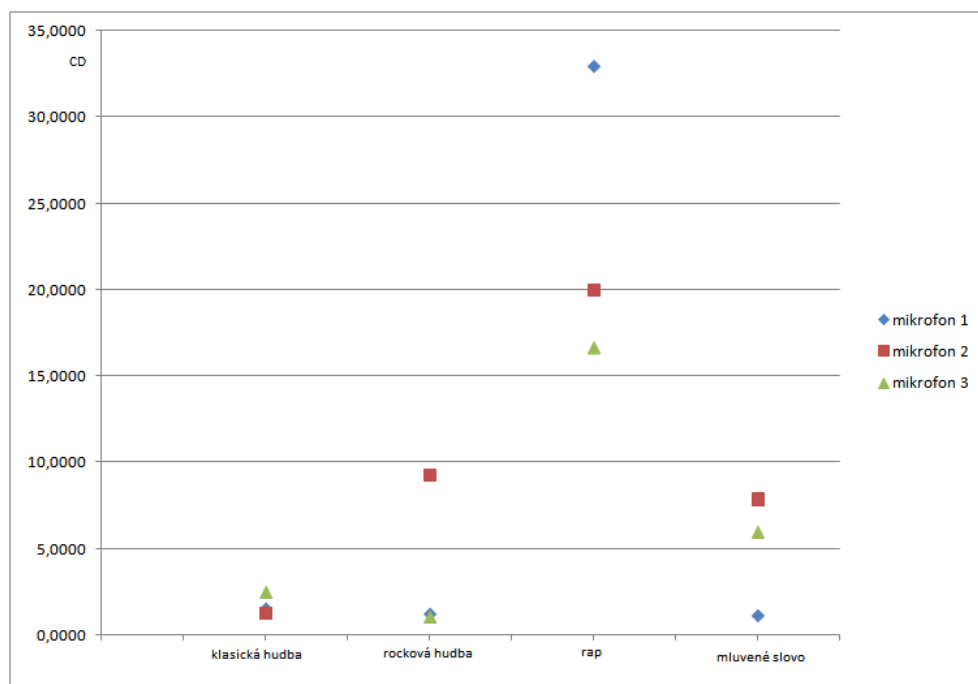
7.2 Vyhodnocení reálných dat

Reálná data byla vyhodnocována pouze CD srovnáním s originální nahrávkou. Celkem byly připraveny čtyři nahrávky a každou zaznamenaly 3 mikrofony. Výsledky jsou ukázány v tabulce 7.13. Výsledky potvrzují, že hudbu obsahující rap zvládá metoda nejhůře. Naopak se nepotvrdilo, že by vzdálenost od zdroje užitečného signálu měla větší vliv. Ovšem blízkost ke zdroji rušení se jeví jako negativní. V případě rockové hudby a mluveného slova měla metoda nejhorší výsledky u druhého mikrofону, který byl zdroji rušení nejbližší.

Ve srovnání testování na reálných a simulovaných datech se nezdá, že by si program vedl na reálných datech výrazně hůře než na simulovaných. Ovšem u reálných dat nebylo rušení zdaleka tak silné, jako v případě některých simulovaných dat. Podle poslechu byla většina rušivého signálu utlumená, ovšem na některých nahrávkách utlumení patrné nebylo téměř znatelné. Graf 7.14 zobrazuje výsledky CD v závislosti na typu rušení u všech tří mikrofónů.

| rušení: | mikrofon 1 | mikrofon 2 | mikrofon 3 |
|----------------|------------|------------|------------|
| klasická hudba | 1,4960 | 1,2792 | 2,4675 |
| rocková hudba | 1,1904 | 9,2580 | 1,0303 |
| rap | 32,9490 | 19,9700 | 16,6300 |
| mluvené slovo | 1,0987 | 7,8702 | 5,9451 |

Obrázek 7.13: Vypočítaná CD na reálných datech.



Obrázek 7.14: Graf výsledků CD reálných dat.

Kapitola 8

Závěr

Tato práce se zabývá odstraňováním rušení z nahrávky v takovém případě, kdy je rušivý signál známý, nebo zjistitelný z jiného zdroje. Zaměřuje se na případy, kdy je nahrávka s užitečným zvukem pořízena nahrávacím zařízením v uzavřené místnosti a rušena jiným zdrojem zvuku, který je možné zjistit. Takovým rušením může být například rádio nebo televize. Pro odstranění takového rušení je nejprve nutné analyzovat, jak bylo známé rušení ovlivněno okolním prostředím, nahrávací technikou, případně i zdrojem samotného rušivého zvuku. Určit toto ovlivnění ovšem můžeme pouze na základě nahraného a známého signálu.

Po seznámení se s problematikou šíření zvuku v uzavřeném prostoru a hledání impulsní odezvy místnosti (kapitola 2) a vytvoření testovacích dat (kapitola 3) jsem analyzoval několik technik pro řešení problému. Zjistil jsem, proč některé techniky nevedou ke splnění cíle a jakými způsoby úspěchu dosáhnout (kapitola 4). Na základě toho jsem navrhnul metodu (kapitola 5), kterou jsem následně implementoval v prostředí MATLAB (kapitola 6). Tuto implementaci jsem nakonec otestoval a vyhodnotil (kapitola 7).

Testování implementované metody ukázalo rozdílné výsledky. Na simulovaných datech byla metoda velmi úspěšná, především při rušení mluveným projevem nebo klasickou hudbou. Horší výsledky měla metoda při rušení užitečného signálu rapem, ovšem pouze když byl v nahrávce výrazně slyšitelnější, než mluvený projev užitečného signálu.

Na reálných datech ukázala metoda výsledky horší. Rušení nebylo tak významné, jako při některých simulovaných datech, přitom né všechny rušivé nahrávky se povedlo slyšitelně vylepšit. Na druhou stranu, u většiny testovacích dat ke znatelnému utlumení známého rušivého signálu došlo. Výsledky však nebyly příliš konzistentní a pro lepší závěry by bylo potřeba větší množství testovacích dat.

Na všech nahrávkách však bylo zřetelné praskání. A to i v případech, kdy byly testovány zcela nerušené nahrávky. Toto praskání přisuzuji práci s rámci při odečítání signálu a další pokusy například s délkou rámců nebo výběrem okenní funkce by mohlo vést k jeho utlumení, možná i úplné eliminaci.

Jako další pokračování práce bych navrhoval upravit metodu tak, aby nepracovala s celými signály, pouze s jejich částmi o délce několika vteřin. Metoda by tak dokázala reagovat na změny v prostředí a dávat lepší výsledky. Zlepšení výsledků by mohla dosáhnout i zavržená metoda odhadu impulsní odezvy na tichých rámcích. Byť v této práci k úspěchu nevedla, při přesnějším odhadu úseků obsahujících pouze rušivý signál by mohla být přece jen nápomocná.

Jinou možností dalšího vývoje práce je implementace metody v jiném programovacím jazyce a vytvoření aplikace s grafickým uživatelským rozhraním. Tato aplikace by mohla být doplněna o další techniky odšumování, případně modulace zvuku. Také by mohla umož-

nit uživateli měnit některé parametry algoritmu, čímž by poskytla širší možnosti a větší univerzálnost.

Literatura

- [1] Brigham, E.: *The Fast Fourier Transform and Its Applications*. Prentice Hall, 1988.
- [2] Elorza, D. O.: *Room acoustics modeling using the ray-tracing method: implementation and evaluation*. University of Turku, Department of Physics, 2005.
- [3] Černý, F.: *Simulace šíření zvukové vlny v uzavřeném prostoru: diplomová práce*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2013.
- [4] Černocký, J.: *Zpracování řečových signálů, studijní opora*. FIT VUT v Brně, 2006.
- [5] Černocký, J.: *Signály a systémy, studijní materiály k předmětu*. FIT VUT v Brně, 2010.
- [6] Šimandl, M.: *Identifikace systémů a filtrace. 1. vyd.* 2001, ISBN 80-7082-170-1.
- [7] Jeub, M.: *Aachen Impulse Response (AIR) Database - Version 1.4*. Aachen University, 2011.
URL <http://www.iks.rwth-aachen.de/en/research/tools-downloads/databases/aachen-impulse-response-database/>
- [8] Klapetek, P.; Nečas, D.; Anderson, C.: *Gwyddion User Guide*. 2011.
- [9] Ljung, L.: *Frequency Domain Versus Time Domain Methods in System Identification*. 2007.
URL <https://pdfs.semanticscholar.org/2ef1/49cbb3fcd3c08a690f8c8fd1f3af8ccb44be.pdf>
- [10] Ljung, L.: *State of the Art in Linear System Identification: Time and Frequency Domain Methods*. 2007.
URL <http://liu.diva-portal.org/smash/get/diva2:316505/FULLTEXT01.pdf>
- [11] *MATLAB dokumentace*. [Online; navštíveno 25.04.2018].
URL https://uk.mathworks.com/help/matlab/learn_matlab/product-description.html
- [12] Žmolíková, K.: *osobní konzultace*. FIT VUT v Brně, 2018.
- [13] Přívara, J.: *Klasifikace objektů s použitím radaru, bakalářská práce*. FIT VUT v Brně.
- [14] Richards, M. A.: *Fundamentals of Radar Signal Processing, 1.vyd.* 2005.
- [15] Smith, S.: *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Pub., 1997.

- [16] Välimäki, V.; Parker, J.; Savioja, L.; aj.: Fifty Years of Artificial Reverberation. ročník 20, 2012.
- [17] Zach, P.: *Metodika sledování a hodnocení počítačové sítě podniku*. Brno, 2015.